韓國開發硏究
제32권 제2호(통권 제107호)

직업훈련의 임금효과 분석: 「경제활동인구조사」를 중심으로

유 경 준
(한국개발연구원 선임연구위원)
강 창 희
(중앙대학교 경제학과 부교수)

The Impacts of Vocational Training on Earnings in Korea: Evidence from the Economically Active Population Survey

Gyeongjoon Yoo
(Senior Research Fellow, Korea Development Institute)
Changhui Kang
(Associate Professor, Department of Economics, Chung-ang University)

* 유경준: 제1저자, 강창희: 교신저자

** 본 연구는 2010년도 중앙대학교 학술연구비의 지원을 받아 수행되었다. 본 연구에 사용된 패널 자료를 제공하여 준 통계청 관계자분들과 자료의 정리와 지리한 전산처리를 성실하게 수행하여 준 KDI의 최바울 주임연구원과 권태구 연구원, 김상신 연구원에게 감사의 뜻을 전한다.

*** 유경준: (e-mail) yoogj@kdi.re.kr, (address) Korea Development Institute, 49 Hoegiro, Dongdaemun-gu, Seoul 130-740, Korea
강창희: (e-mail) ckang@cau.ac.kr, (address) Chung-Ang University, 211 Heukseok-dong, Dongjak-gu, Seoul, Korea

• Key Word: 직업훈련(Vocational Training), 임금효과(Wage Effect), 성향점수 매칭법(Propensity-score Matching)
• JEL Code: C0, C3, C8
• Received: 2009. 10. 23 • Referee Process Started: 2009. 10. 23
• Referee Reports Completed: 2010. 2. 19

ABSTRACT

This paper examines whether and how much vocational training raises an individual's earnings in Korea, using the Economically Active Population Survey. To overcome endogeneity of training, we apply fixed-effects and propensity-score matching (PSM) methods. Fixed-effects (PSM) results suggest that work-related training received in the previous one year increases a worker's monthly earnings by 2.6 to 4.7 (7.5 to 9.8) percent. Taken altogether, work-related training enhances a worker's earnings by a minimum of 2.6 and a maximum of 9.8 percent in Korea.

본 논문은 우리나라에서 직업능력 향상을 위한 교육훈련이 개인의 임금에 어떤 영향을 미치는지를 「경제활동인구조사」 자료를 통해 분석한다. 교육훈련의 내생성을 통제하기 위하여 실증분석 방법으로 (1) 고정효과 추정법과 (2) 성향점수 매칭법을 적용한다. 고정효과 추정법(매칭법)의 결과에 따르면, 지난 1년 동안에 이수한 교육훈련으로 인하여 개인의 월평균 임금수준은 평균 2.6～4.7%(7.5～9.8%) 정도 상승한다. 상이한 두 가지 추정방법의 분석 결과를 종합하면, 우리나라에서 직업능력 향상을 위한 교육훈련의 임금 상승효과는 평균 2.6~9.8% 수준인 것으로 추정된다.

Ⅰ. 서 론

노동시장에 참여하는 개인들은 학교와 학교 이외의 직업훈련기관들을 이용하여 전 생애에 걸쳐서 인적자본을 형성한다. Heckman et al.(1998)의 연구 결과에 따르면, 개인의 생애 인적자본(lifetime human capital) 중 절반 이상은 학교를 졸업한 이후에 형성된다. 학교 졸업 이후에 형성되는 인적자본의 중요성에도 불구하고, 대부분의 경제학 연구들은 주로 학교를 통한 인적자본 형성과 그 효과에 초점을 맞추어 왔다(Card[2000]). 상대적으로 작은 규모이기는 하지만, 서구 노동경제학계에서는 학교 졸업 후의 직업훈련이 개인의 생산성과 임금에 미치는 효과에 관한 연구가 꾸준히 진행되어 왔다. 서구에서 진행된 직업훈련의 효과에 관한 최근의 연구들은 Bassanini et al.(2005, 4장), Blundell et al.(1999) 및 Leuven and Oosterbeek(2008)와 이들 논문에서 인용된 여러 연구들을 통하여 확인할 수 있다.

서구와 비교하면, 우리나라에서는 직업훈련이 개인에게 미치는 효과에 대한 연구가 그리 활발한 상태는 아니다. 최근에 경제학 저널에 발표된 연구들로 범위를 제한하면, 강순희·노홍성(2000), 김안국(2002), 이상호(2005) 정도에 불과하다. 위의 논문들은 우리나라 직업훈련의 효과 연구에서 가지는 선도적인 의미에도 불구하고, 직업훈련의 인과효과를 정확하게 보여주지 못하고 있으며, 몇 가지 한계를 가지고 있다.

먼저, 강순희·노홍성(2000)은 「한국노동패널조사」의 1차연도인 1998년 횡단면 자료를 사용하여 조사시점 이전에 받은 직업훈련이 조사시점에서의 취업 및 임금에 긍정적인 효과가 있음을 발견하였다. 이 분석에서는 직업훈련의 내생성을 고려하기 위하여 Heckman의 선택편의 교정모형을 사용하고 있는데, 1단계의 훈련참여 방정식에는 포함되고 2단계의 주 방정식에서는 제외되는 식별변수(identifying variables)가 논문에 명시적으로 제시되지 않는다. 분석에서 추정된 효과가 각 단계에서 사용된 함수들의 비선형성에 의하여 식별되는 점을 감안할 때 논문의 실증결과를 전적으로 신뢰하기에는 한계가 있다. 더구나, 자료구조상의 문제로 인하여, 개인이 직업훈련을 받은 정확한 시점을 알 수 없어서, 이 효과가 진정한 직업훈련의 효과인지 그렇지 않은지를 구분하기 어렵다. 예를 들어, 조사시점 약 5년 전에 직업훈련을 이수한 개인은 1998년 조사 당시 직업훈련 경험자로 분류되지만 실제 훈련의 효과는 소멸되었을 가능성이 있고, 이때 추정된 직업훈련의 효과는 측정오차(measurement errors)로 인하여 오염될 수 있기 때문이다.

김안국(2002)은 「한국노동패널조사」의 1998년부터 2000년의 3개년 자료를 이용하여 직업훈련의 효과를 고정효과모형을 통하여 추정하였다. 고정효과모형은 Heckman의 선택편의 교정모형보다 약한 가정에 기반하고 있다는 점에서 직업훈련의 효과 분석에 보다 유용한 모형으로 판단된다. 이 연구에서는 교육훈련이 개인의 임금을 평균 5% 정도 상승시키지만, 그 크기가 통계적으로 유의하지는 않다고 보고한다. 이 연구에서도 강순희·노홍성(2000)과 마찬가지로 「한국노동패널조사」의 1998년 자료에서 관측된 '직업훈련 여부'를 중요 설명변수로 사용하고 있기 때문에 직업훈련시점이 명시되지 않는 한계가 있다(1999년 조사에서는 조사 이전 1년간의 직업훈련 이수 여부를 묻고 있기 때문에 훈련시점의 문제가 상대적으로 작다). 강순희·노홍성(2000)과 동일하게, 직업훈련 여부의 측정오차로부터 훈련의 추정효과가 영에 가까워지는 편의(attenuation bias)가 발생할 수 있다.

이상호(2005)는 직업훈련시점에 대한 측정오차가 상대적으로 작은 「한국노동패널조사」의 2003년과 2004년 자료와 고정효과모형을 이용하여 직업훈련의 효과를 추정하였다. 그의 연구 결과에 의하면, 직업훈련의 임금 상승효과는 1.4%에 불과하고 통계적으로 유의하지 않다. 이 논문에서는 효과의 크기에 대해서만 간략하게 설명되고 상세한 분석이 진행되지는 않았기 때문에 그 분석 결과를 보다 상세히 검토할 필요성이 제기된다.

본 연구에서는 우리나라에서 개인의 직업능력 향상을 위한 교육훈련이 임금에 미치는 효과를 「한국노동패널조사」가 아닌 「경제활동인구조사」를 통해 구축되는 패널 자료를 이용하여 분석한다. 동일한 주제를 상이한 자료를 이용하여 분석함으로써 훈련의 효과에 대한 보다 풍부한 실증 결과를 얻을 수 있을 것으로 기대된다. 위와 같은 자료상의 차이점 이외에도 본 연구는 다음과 같은 세 가지 부분에서 이전의 연구들과 구분된다.

첫째, 앞서 설명한 바와 같이, 강순희·노홍성(2000)과 김안국(2002)의 연구에서는 1998년의 「한국노동패널조사」 자료가 사용되었기 때문에 직업훈련 이수 시점에 대하여 명확히 판단하기 힘들다. 즉, 1998년 이전에 이미 직업훈련을 받았고 그 효과가 이미 소멸되었더라도 1998년 조사에서는 직업훈련 이수자로 분류된다. 이 경우 1998년에 조사된 직업훈련 이수 여부 설명변수는 측정오차로 오염되어 훈련의 효과가 유의미하지 않게 추정될 가능성이 존재한다. 반면에, 본 연구에서 사용되는 「경제활동인구조사」에서는 '지난 1년간 직업능력개발을 위한 교육훈련을 받은 경험이 있는지'가 조사되고 있어 훈련 이수 여부에 대한 측정오차가 상대적으로 작을 가능성이 높다. 보

다 정확한 훈련시점을 식별할 수 있는 분석 자료를 사용함으로써 본 연구는 선행연구에서 추정된 것보다 우수한 훈련효과의 추정치를 추출할 수 있을 것으로 기대된다.

둘째, 직업훈련의 효과 추정에서는 횡단면 자료를 이용하는 Heckman의 선택편의 교정모형보다도 패널 자료를 이용하는 고정효과모형이 일반적으로 선호된다. 직업훈련의 이수 여부에 존재하는 내생성을 통제하기에 후자의 모형이 보다 유용하기 때문이다. 이 점을 고려하여 김안국(2002)과 이상호(2005)는 고정효과모형을 직업훈련효과 추정에 적용하였다. 그러나 개인들의 훈련 이수 결정에서 시점간의 임금 변동이 중요한 역할을 한다면, 고정효과모형은 훈련의 효과에 대하여 상향편의된(upward-biased) 추정치를 제공할 가능성이 있다. 예를 들어, 전년도에 임금이 상대적으로 많이 떨어진 개인들이 체계적으로 올해 직업훈련을 더 많이 받는다면 고정효과모형의 추정치는 실제 훈련의 효과를 과대평가할 가능성이 있다. 본 논문에서는 고정효과모형과 더불어 임금수준을 종속변수로 사용하는 비모수적 성향점수 매칭법(propensity-score matching methods)을 적용함으로써 추정방법에 따라 훈련의 효과가 상이하게 나타나는지 확인한다. 양 방법을 통해 구한 추정치가 유사할 경우 추정 결과의 신뢰도는 높아질 수 있다.

셋째, 앞서의 연구들은 주로 1999년 이전 또는 2003/04년에 이수한 훈련에 대한 임금효과를 추정하였다. 본 연구는 2008년과 2009년에 이수된 직업훈련에 대한 효과를 추정함으로써 가장 최근의 훈련효과에 대한 정보를 제공해 준다.

본 논문의 고정효과 패널 추정법의 결과에 따르면, 지난 1년 동안에 이수한 교육훈련으로 인하여 개인의 월평균 임금수준은 평균 2.6~4.7% 정도 상승하는 것으로 추정된다. 한편, 매칭법의 추정 결과에 의하면, 직업능력 향상을 위한 교육훈련은 개인의 월평균 임금을 평균 약 7.5~9.8% 정도 상승시킨다. 그리고 고정효과 및 매칭의 추정치들은 5%의 유의수준에서 대부분 유의하다. 두 가지의 추정방법을 통한 분석 결과를 종합하면, 우리나라에서 직업능력 향상을 위한 교육훈련의 임금 상승효과는 평균 2.6~9.8% 수준인 것으로 추정된다.

본 논문은 다음과 같은 순서로 전개된다. 제Ⅱ장에서는 자료를 분석하는 데 사용되는 고정효과 통계모형과 성향점수 매칭법이 설명된다. 제Ⅲ장에서는 분석자료인 「경제활동인구조사」가 소개되고, 제Ⅳ장에서는 교육훈련의 효과에 대한 실증분석 결과가 설명된다. 마지막으로 제Ⅴ장에서는 논문의 결론이 제시된다.

II. 계량분석모형

1. 고정효과(Fixed Effects) 모형

본 논문의 실증분석에 사용되는 임금 및 개인의 특성에 관한 자료는 통계청이 작성하는 「경제활동인구조사」(이하 경활조사)의 2007년 3월, 2008년 3월과 2009년 3월의 본조사 및 부가조사 자료이다. 경활조사는 본래 월별로 반복되는 횡단면(repeated cross-section) 조사이지만, 조사에 진입한 개인이 일정한 기간 동안 표본에 머무는 특성을 이용하여 지속관찰 자료를 구축할 수 있다. 본 논문에서는 (1) 2007년 3월과 2008년 3월의 2개년 조사를 통하여 구축되는 10,790명의 개인에 대한 패널 자료(패널 1), (2) 2008년 3월과 2009년 3월의 2개년 조사를 통하여 구축되는 11,401명의 개인에 대한 패널 자료(패널 2), 그리고 (3) 2007년 3월, 2008년 3월 및 2009년 3월의 3개년 조사를 통하여 구축되는 3,962명의 개인에 대한 패널 자료(패널 3)를 이용하여 교육훈련의 임금효과를 추정한다.

개인의 임금방정식에서 직업훈련의 내생성을 통제하는 분석방법으로서 우리는 먼저 패널 자료에 기반한 다음의 고정효과모형을 사용한다.

$$y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + c_i + \epsilon_{it} \qquad (1)$$

식 (1)에서 i는 개인을, t는 연도를 의미하며, $y_{it} \equiv \ln(w_{it}) - \ln(w_{it-1})$로서 $(t-1)$시점에서 t시점 사이 개인 i의 월평균 임금(w)의 증가율,[1] Z_{it}는 개인 i가 t시점 이전의 지난 1년 동안 교육훈련을 받았는지를 나타내는 더미변수, X_{it}는 개인이나 사업장의 특성을 통제하기 위한 독립변수 벡터,[2] c_i는 시간의 흐름에 따라 변하지 않으면서 관측되지 않는 개인의 특성, 마지막으로 ϵ_{it}는 임의의 교란항을 나타낸다.

횡단면 자료를 이용한 직업훈련의 효과 분석에서는 종속변수로서 '$\ln(w_{it}) - \ln(w_{it-1})$'의 차분(difference) 변수 대신에 '$\ln(w_{it})$'의 수준(level) 변수가 통상적으로 이용된다. 그러나 이 경우에는 개인의 관찰되지 않는 특성(c_i)과 교육훈련더미

1) 종속변수로서 월평균 임금 대신에 시간당 임금을 사용하는 대안적인 모형을 생각해 볼 수 있다. 그러나 시간당 임금에 대한 추정 결과도 본 논문에 보고된 월평균 임금에 대한 결과와 질적으로 크게 다르지 않다. <Appendix Table 1>에는 시간당 임금의 차분을 종속변수로 사용한 고정효과모형의 추정 결과가 참고로 제시되어 있다.

2) 설명변수로서 성별, 가구주 여부, 기혼 여부, 연령, 도시거주 여부, 근속연수, 교육연수, 비정규직 여부, 직업(비교집단: 준전문직, 사무직), 산업(비교집단: 제조업), 사업장규모(비교집단: 5인 미만)가 포함된다.

(Z_{it}) 간에 존재하는 상관관계로 인하여 훈련효과에 대한 OLS 추정량에 편의가 발생할 가능성이 높다. 예컨대, 생산성 또는 업무에 관한 성취욕이 높은 근로자가 그것이 낮은 근로자보다 평균적으로 교육훈련을 이수할 가능성이 높다면 γ의 OLS 추정치는 과대추정(over-estimated)될 수 있다. 반대로, 기업에서 생산성이나 지적인 능력이 떨어지는 근로자에 대하여 체계적으로 교육훈련을 더 시킨다면, γ의 OLS 추정치는 과소추정(under-estimated)될 수 있다. 본 논문에서 사용되는 고정효과 추정방법에서는 Z_{it}의 내생성을 통제하기 위하여 종속변수로서 '$\ln(w_{it})$' 대신에 '$\ln(w_{it}) - \ln(w_{it-1})$'을 이용한다.

먼저, '단순한' 고정효과 추정법으로서, 본 논문은 모든 i에 대하여 '$c_i = 0$'이라고 가정하고 '패널 1'과 '패널 2' 각각에 대하여 식 (1)을 OLS 방법을 적용하여 추정한다. 종속변수로서 '$\ln(w_{it}) - \ln(w_{it-1})$'을 사용하는 OLS 추정법은 종속변수로서 '$\ln(w_{it})$'를 사용하는 횡단면 OLS 추정법에 비하여 Z_{it}의 내생성이 상대적으로 작다고 알려져 있다. 왜냐하면, '$\ln(w_{it-1})$'을 통하여 개인의 관측되지 않는 특성들이 상당 부분 통제될 수 있기 때문이다. 그러나 '패널 1'과 '패널 2'를 이용한 분석에서 Z_{it}의 내생성이 상대적으로 작다고 하더라도 그것이 완전히 제거되었다고 판단하기는 어렵다. 왜냐하면, '$c_i = 0$'이 성립하지 않는 경우 이 추정법을 통해서는 c_i의 영향이 완벽하게 통제되었다고 생각할 수 없기 때문이다.

이에 대한 해결책으로서 본 논문은 3년간의 개인 패널인 '패널 3'을 이용하고 c_i를 고정효과(fixed effects)로 처리하는 '확장된' 고정효과 추정법을 적용한다. '패널 3'을 이용한 고정효과 분석법에서는 원래의 변수 관측값에 대한 within transformation[3]을 통하여 내생성을 유발하는 c_i가 자연스럽게 제거된다. '패널 3'을 이용한 고정효과 분석법이 가지는 통계적인 우월성에도 불구하고, 이 방법은 3년간의 패널자료에 대해서만 적용될 수 있기 때문에 소규모 샘플로부터 발생하는 편의(small-sample bias)를 통제하기 어렵다. '패널 3'이 대표성을 상실한다면 이로부터 구해진 직업훈련의 추정치 또한 신뢰하기 힘들 것이다.

샘플의 대표성과 관측치 수의 관점에서 볼 때, '패널 1'과 '패널 2'를 이용한 추정법은 '$c_i = 0$'이라는 제약적인 가정을 부과하면서 동시에 '패널 3'의 약점을 보완하는 실증방법으로 이해된다. 본 논

3) within transformation: $(y_{it} - \overline{y_i}) = \sum_{j=1}^{J} \beta_j (x_{it}^j - \overline{x_i^j}) + \gamma(Z_{it} - \overline{Z_i}) + (\epsilon_{it} - \overline{\epsilon_i})$ (2)

여기서 $\overline{y_i} = \frac{1}{T_i}\sum_t^{T_i} y_{it}$, $\overline{x_i^j} = \frac{1}{T_i}\sum_t^{T_i} x_{it}^j$, $\overline{Z_i} = \frac{1}{T_i}\sum_t^{T_i} Z_{it}$, and $\overline{\epsilon_i} = \frac{1}{T_i}\sum_t^{T_i} \epsilon_{it}$.

문은 '패널 1', '패널 2' 및 '패널 3'을 이용하는 각각의 추정방법들에 서로 다른 장점과 단점이 존재하는 점을 고려하여 각 샘플로부터 구해진 직업훈련효과의 추정치들에 동일한 가중치를 부여하여 해석한다. 세 가지 분석 결과의 추정치들이 자료 및 모형 가정상의 차이에도 불구하고 꽤 좁은 범위 안에 모여 있는 경우, 우리는 각 추정치들이 신뢰할 만하다고 판단할 수 있을 것이다.

2. 성향점수 매칭법 (Propensity-score matching)

앞서 언급한 바와 같이, 고정효과를 이용한 추정모형은 내생성을 유발하는 개인효과를 차분을 통하여 제거하는 유용한 방법을 제공하지만, 교육훈련의 효과에 대하여 상향편의된(upward-biased) 추정치를 제공할 가능성이 있다. 이 가능성은 처치효과(treatment effects)를 다루는 문헌에서는 'Ashenfelter's dip'이라고 불린다. 직업훈련을 받는 사람들의 성향을 조사하면, 일반적으로 (다른 조건이 동일할 때) 훈련 전에 임금이 상대적으로 많이 떨어진 개인들이 체계적으로 직업훈련을 더 많이 받는다. 이 경우 고정효과모형을 이용하면 $Cov(Z_{it}, c_i) \neq 0$의 가능성은 통제할 수 있으나, $Cov(Z_{it}, \epsilon_{it-1}) < 0$의 가능성 때문에 훈련의 효과가 과대추정될 수 있다. 그리고 일반적으로 차분을 이용한 추정에서는 원래의 수준값을 이용하는 추정에서보다 Z_{it}에 내재한 측정오차의 문제가 심각해져서, 훈련의 효과가 존재하지 않는 것처럼 나타날 수 있다(attenuation bias). 이러한 문제를 보완하기 위하여 본 논문에서는 고정효과모형과 더불어 변수들의 원래 수준값을 이용하는 성향점수 매칭법(propensity-score matching methods)을 적용하고, 두 개의 추정방법 간에 훈련효과가 상이하게 나타나는지 검토한다.

성향점수 매칭법은 다음과 같이 전개된다. 교육훈련의 순수한 효과는 훈련 참여자들이 훈련에 참여한 이후 얻은 노동성과와, 만약 동일한 참여자가 훈련에 참여하지 않았더라면 얻었을 성과 간의 차이라고 정의할 수 있다.

$$\tau = E\{Y_i^1 - Y_i^0 | Z_i = 1\} = E\{Y_i^1 | Z_i = 1\} - E\{Y_i^0 | Z_i = 1\} \quad (2)$$

여기서 Z_i는 개인 i가 지난 1년 동안 교육훈련을 받았는지를 나타내는 더미변수이고, $Y_i^j\,(j = 0,1)$는 j의 상태에서 발생하는 잠재적인 성과(potential outcome)이다. 앞 절에서와 달리, 본 절의 Y는 수준변수인 $\ln(w)$로 정의된다. τ는 흔히 처치수혜자($Z_i = 1$)에 대한 처치의 평균효과(Average Treatment Effect on the Treated: ATT)라고 불린다. 교육훈련의 관점에서

이 식을 해석하면, $E\{Y_i^1|Z_i=1\}$은 훈련 이수자의 훈련 이후 노동시장 성과(즉, 임금)를 나타내고, $E\{Y_i^0|Z_i=1\}$은 실제 훈련 이수자가 훈련을 이수하지 않았다면 얻게 되었을 가상적 상황에서의 노동시장 성과(가상적 대응치, counterfactual)를 의미한다.

일반적인 정책평가에 있어 기본적인 문제는, $E\{Y_i^1|Z_i=1\}$은 자료에서 쉽게 추정될 수 있는 반면에 $E\{Y_i^0|Z_i=1\}$은 자료에서 직접적으로 관측되지 않는다는 점이다. 매칭기법은 관측되지 않는 $E\{Y_i^0|Z_i=1\}$을 추정하기 위하여 처치 참여자와 동일한 관측 특성(X)을 갖는 비참여자의 성과를 이용하는 통계방법이다. 관측 가능한 특성에 기반하여 참여자와 비참여자를 대응시키기 때문에, 매칭법은 다음과 같은 조건부 독립성 가정(Conditional Independence Assumption: CIA)에 기초하고 있다.

$$(Y^0, Y^1) \perp Z \mid X \qquad (3)$$

이 가정은 관측되는 특성(X)이 주어져 있을 때 한 개인의 처치 여부(Z)는 잠재적인 성과(Y^0, Y^1)와는 독립적이라는 것이다. 이 것으로부터 '$E(Y_i^0|Z_i=1, X_i) = E(Y_i^0|Z_i=0, X_i)$'이 성립한다. 즉, X_i라는 관측 특성을 갖는 처치 참여자가 미참여 시 받게 되는 평균 가상성과 $[E(Y_i^0|Z_i=1, X_i)]$는 참여자와 동일한 관측 특성을 갖는 처방 미참여자의 관측된 평균 성과$[E(Y_i^0|Z_i=0, X_i)]$에 의하여 추정된다. 참여자가 여러 가지 다양한 관측 특성(X)을 가질 수 있으므로, 실제 처방 참여자가 미참여 시 받게 되는 평균 가상성과는 X에 대하여 $E(Y_i^0|Z_i=0, X_i)$를 가중평균함으로써 구해진다. 즉, '$E\{Y_i^0|Z_i=1\} = E_X\{E(Y_i^0|Z_i=0, X)|Z_i=1\}$'의 관계가 성립한다. 결국 교육훈련의 평균효과는 '$\hat{\tau} = \hat{E}\{Y_{1i}|Z_i=1\} - \widehat{E_X}\{\hat{E}(Y_{0i}|Z_i=\}\{0, X)|Z_i=1\}$의 식을 통하여 추정된다.

그러나 실제 이 방법을 적용할 때에는 X가 한 변수가 아닌 여러 개의 변수로 구성될 수 있다. 이 경우 모든 변수의 조합에 대하여 동일한 값을 갖는 두 명의 짝을 찾는 것이 현실적으로 어렵거나 불가능할 수 있다. Rosenbaum and Rubin (1983)은 성향점수 $\Pr(D=1|X)=p(x)$를 이용한 매칭을 통해서도 X를 이용하는 매칭과 동일한 인과효과를 추정할 수 있음을 증명하여, X의 다차원성의 문제(curse of dimensionality)를 해소하였다. Rosenbaum and Rubin(1983)은 가정 1과 2가 추가될 경우 성향점수를 이용한 매칭을 통해서 선택편의가 없는 처치효과의 추정치를 구할 수 있음을 보였다.

가정 1. $0 < p(x) < 1$

가정 2. $Z \perp X \mid p(x)$

위의 가정들이 성립할 경우 처치집단의 한 구성원과 성향점수가 동일한(보다 정확하게는 근사한) 비교집단의 구성원을 짝지어서 그 성과를 차분하여 처치효과를 추정하며, 이 추정치는 무작위 실험평가와 동일하게 선택편의가 존재하지 않는다. 성향점수를 이용한 매칭에서는 다음의 식을 통하여 훈련의 효과가 추정된다.

$$\hat{\tau} = \frac{1}{n_1} \sum_{i \in (I_1 \cap S_p)} \left[Y_i^1 - \hat{E}(Y_i^0 | Z_i = 1, p_i) \right] =$$

$$= \frac{1}{n_1} \sum_{i \in (I_1 \cap S_p)} \left[Y_i^1 - \sum_{j \in I_0} W(i,j)\, Y_j^0 \right]$$

여기서 I_1은 훈련 참여자들의 집합, I_0는 비참여자들의 집합, S_P는 가정 1에서 제시되는 바와 같이 참여자와 비참여자가 동시에 존재하여 '$0 < p(x) < 1$'이 성립하는 공통 영역(region of common support), N_1은 $I_1 \cap S_P$ 집합에 있는 관측치의 숫자이고, $W(i,j)$는 p_i와 p_j 간의 거리에 대해서 부여되는 가중치를 표시한다. 성향점수 매칭에서는 $p(x)$가 연속변수이기 때문에 정확히 일치하는 성향점수를 갖는 두 개의 관측치를 찾을 확률은 0이어서, $W(i,j)$의 가중치를 부여하는 방법에 따라 다양한 매칭기법이 사용된다.

첫 번째 방법인 stratification matching에서는 성향점수의 변화 범위를 구간으로 분리한다. 이때 각각의 구간에서는 처치·비교 집단이 평균적으로 동일한 성향점수를 갖도록 구간을 분리한다. 처치집단과 비교집단이 모두 존재하는 구간마다 각 집단의 평균 성과의 차이(ATT)를 계산하고, 각 블록에서 얻어진 ATT의 가중평균을 구하여 최종적인 ATT를 구한다. 가중치는 블록 간 처치집단 분포에 의해 결정되며, 처치집단과 비교집단 중 어느 하나가 없는 블록에 대해서는 ATT 계산에 포함되지 않으므로 관측치가 삭제된다.

두 번째 방법인 kernel matching은 모든 처치집단 내 개별 관측치를 비교집단의 모든 관측치와 짝 지우는 방법이다. 각 짝의 가중치는 처치집단의 성향점수와 비교집단의 성향점수 간의 거리에 반비례하도록 설정한다.

세 번째 방법인 nearest neighbor matching은 한 처치 관측치를 성향점수가 가장 가까운 비교집단의 관측치 하나와 짝 지우는 방법이다. 이 방법을 통해서는 각 처방 관측치에 대하여 매칭되는 짝을 항상 찾을 수 있으나 둘 사이의 성향점수가 크게 차이 날 수도 있는 단점이 있다.

네 번째 방법인 radius matching은 각 처치 관측치에 대해 미리 정해 놓은 성향점수 범위(반지름) 내에 포함되는 성향점수를 가진 관측치만을 비교 대상으로 선정하는 방법이다. 반지름이 작으면 비교대상의 질이 우수하나 관측치를 찾지 못

할 수도 있다.

위 매칭기법들은 어느 한 기법이 나머지 다른 기법보다 통계적으로 우월하지 않으며, 각기 상이한 장단점을 가지고 있다(Smith and Todd[2005]). 따라서 본고에서는 위에 소개한 4가지 매칭 방법을 모두 사용하여 ATT를 추정하고, 추정량들의 전반적인 추세를 관찰하고자 한다.

Ⅲ. 분석 자료

본고의 분석에 사용되는 자료는 통계청의 「경제활동인구조사」(경활) 2007년 3월, 2008년 3월 및 2009년 3월의 본조사 및 부가조사 자료이다. 2007년 8월 이전의 경활 부가조사에서는 '지난 1년간 교육훈련을 받은 경험이 있는지'를 질문하고 있고, 위의 교육훈련에는 취미·교양 목적의 교육훈련이 포함될 수 있다. 그러나 2007년 8월부터는 '지난 1년간 직업능력개발을 위한 교육훈련을 받은 경험이 있는지'를 조사하고 있어, 본 연구에서 관심을 가지고 있는 직업능력개발을 위한 직업훈련 실시 여부가 적절히 관측된다. 차분을 이용하는 아래의 실증분석에서는 이전 연도의 임금이 종속변수에 포함되기 때문에, 2008년 3월 및 2009년 3월의 자료와 더불어 2007년 3월 자료의 임금 정보를 추가적으로 이용한다. 2009년에 직업 및 산업 분류가 각각 6차 직업분류와 9차 산업분류로 변경되었는바, 본고에서는 이를 고려하여 직업 및 산업 분류상의 일관성을 유지하였다.

<Table 1>에는 실증분석에 사용되는 자료와 주요 변수들의 기술 통계량이 보고되어 있다. 표의 (1)열에는 2007년 3월과 2008년 3월의 경활조사를 연결한 자료인 '패널 1'에 대한 기술 통계량이 제시되어 있고, (2)열에는 2008년 3월과 2009년 3월의 조사를 연결한 자료인 '패널 2'에 대한 기술 통계량이 보고되어 있다. (3)열에는 2007년 3월, 2008년 3월, 2009년 3월의 세 경활조사를 연결한 '패널 3'에 대한 기술 통계량이 제시되어 있다.

표본을 구축한 과정에 대하여 좀 더 자세히 설명하면 다음과 같다.

표본 ①은 2007년 3월과 2008년 3월 양 시점에 모두 임금근로자인 10,790명의 개인들로 구성되고, 표본 ②는 2008년 3월과 2009년 3월 양 시점에 모두 임금근로자인 11,401명의 개인들로 구성된다. 그리고 표본 ③은 2007년 3월, 2008년 3월 및 2009년 3월의 세 시점에 모두 임금근로자인 3,962명의 개인들이고, 2008년 3월 및 2009년 3월의 두 시점을 분석에 이용하기 때문에 표본의 관측치 수는 총 7,924개이다. 훈련 여부에 따라 각 시점 간 임금근로자로서의 지위가 유지되는 패턴을 확인한 결과는 <Table 2>에 제시되어 있다.

〈Table 1〉 Descriptive Statistics of the Samples: Means and Standard Deviations

(Unit: year, %)

		0703~0803	0803~0903	0703~0903
		①	②	③
Number of Observation		10,790	11,401	7,924
Wage	log(wage)	0.03 (0.34)	0.02 (0.33)	0.02 (0.32)
Job Training	Trained	0.27 (0.45)	0.34 (0.47)	0.32 (0.47)
	Untrained	0.73 (0.45)	0.66 (0.47)	0.68 (0.47)
Gender	male	0.57 (0.50)	0.58 (0.49)	0.58 (0.49)
	female	0.43 (0.50)	0.42 (0.49)	0.42 (0.49)
Head of the Family	the head of the family	0.62 (0.49)	0.62 (0.49)	0.63 (0.48)
	other members	0.38 (0.49)	0.38 (0.49)	0.37 (0.48)
Marital Status	married	0.70 (0.46)	0.71 (0.45)	0.72 (0.45)
	single	0.30 (0.46)	0.29 (0.45)	0.28 (0.45)
City · District	City(시)	0.83 (0.38)	0.84 (0.37)	0.84 (0.37)
	District(군)	0.17 (0.38)	0.16 (0.37)	0.16 (0.37)
Age	years	41.96 (11.39)	42.30 (11.23)	43.12 (11.07)
Education	years	12.69 (3.35)	12.84 (3.31)	12.77 (3.38)
Hours Worked	hours	195.3 (47.40)	192.7 (45.96)	194.2 (44.91)
Occupation	management, professional	0.14 (0.34)	0.23 (0.42)	0.19 (0.39)
	service, sales	0.16 (0.36)	0.17 (0.37)	0.15 (0.36)
	other(agriculture, forestry, technic, assembly etc.	0.24 (0.43)	0.23 (0.42)	0.23 (0.42)
	simple labor	0.15 (0.36)	0.16 (0.37)	0.15 (0.36)
	semi-professional, office work	0.31 (0.46)	0.21 (0.41)	0.28 (0.44)
Industry	agriculture, forestry, fishing and mining	0.01 (0.12)	0.01 (0.11)	0.01 (0.11)
	electricity, gas, water and construction	0.08 (0.27)	0.09 (0.29)	0.09 (0.28)
	wholesale · retail sale and accommodation	0.14 (0.37)	0.17 (0.38)	0.16 (0.37)
	transportation and communication	0.06 (0.23)	0.08 (0.27)	0.07 (0.25)
	financial, insurance and business service	0.17 (0.38)	0.16 (0.37)	0.17 (0.37)
	public administration, education and health	0.21 (0.41)	0.22 (0.41)	0.23 (0.42)
	other(other than manufacturing)	0.07 (0.26)	0.06 (0.24)	0.06 (0.24)
	manufacturing	0.23 (0.42)	0.22 (0.41)	0.21 (0.41)
Size	Less than 5	0.26 (0.44)	0.27 (0.44)	0.25 (0.43)
	5～ 29	0.38 (0.48)	0.38 (0.49)	0.38 (0.48)
	30～ 99	0.22 (0.41)	0.21 (0.41)	0.22 (0.42)
	100～299	0.10 (0.30)	0.10 (0.30)	0.11 (0.31)
	More than or equal to 300	0.04 (0.19)	0.04 (0.19)	0.04 (0.21)
Regular work	regular work	0.67 (0.47)	0.69 (0.46)	0.70 (0.46)
	non-regular work	0.33 (0.47)	0.31 (0.46)	0.30 (0.46)
Separation	separated	0.28 (0.45)	0.29 (0.45)	0.26 (0.44)
	continued to work	0.72 (0.45)	0.71 (0.45)	0.74 (0.44)

Note: Wage earners are observed at each point of time of interest. The weights given by Current Population Survey were not used.[4)]

〈Table 2〉 Attrition Rates by the Training Status

		Trained(%)	Untrained(%)	Total(%)
Sample ① (0703 & 0803)	Dropped	55.5	61.2	59.6
	Retained	44.5	38.8	40.4
Sample ② (0803 & 0903)	Dropped	49.5	57.1	55.2
	Retained	50.5	42.9	44.8
Sample ③ (0703 & 0803 & 0903)	Dropped	82.8	86.1	85.2
	Retained	17.2	13.9	14.8

Note: Whether the observant was trained or not is based on the survey question given to the panel sample at the start of the year.

전반적으로 훈련을 받은 근로자의 표본 유지율이 상대적으로 다소 높게 나타나고 있다. 그러나 이러한 차이가 큰 편은 아닌 것으로 판단되어(전체 평균 대비 ± 5% 내외) 본고의 분석에 큰 무리는 없을 것으로 판단된다.[5]

<Table 1>에 따르면, 매년 약 27~34%의 근로자들이 직업능력 향상을 위한 교육훈련에 참여하고 있다.[6] 2007년 3월~2008년 3월 사이에 교육훈련에 참여한 근로자의 비중은 약 27%이고, 2008년 3월~2009년 3월 사이에 교육훈련에 참여한 근로자의 비중은 약 34%로서 후자의 기간에 교육훈련 참여도가 약간 높다. 각 시기별로 약간의 차이가 있으나, 사용되는 자료에서 남성의 비중은 약 57~58%, 가주주의 비중은 약 62~63%, 기혼자의 비중은 약 70~72%, 도시거주자의 비중은 83~84%, 정규직의 비중은 67~70%, 이직자의 비중은 약 26~28% 정도로 분석 자

4) 본고에서 나타나는 임금근로자의 통계분석 결과는 「경제활동인구조사」 결과 도출에 사용되는 가중치를 사용하지 않은 것으로서 임금, 종사상지위, 교육훈련 등에 관한 둘 사이의 단순 비교는 옳지 않을 수 있다.

5) 이러한 차이는 추정치가 편의될 가능성을 남기고 있다. 하나의 가능성은, 훈련을 받은 근로자 중 능력이 더 높은 근로자들이 1년 뒤에 임금근로자로 남는 경향이 더 크다면, 단순한 추정은 훈련의 효과를 과대추정할 소지가 있다는 것이다.

6) 본고에서 사용하는 「경제활동인구조사」상의 직업훈련 변수에는 훈련의 대상, 기간, 내용에 대한 구체적인 변수는 포함하고 있지 않다. 따라서 훈련의 구체적인 종류를 구분한 분석은 다루지 못하고 있다. 또한 다른 직업훈련 자료로서 본고에 비교될 수 있는 자료는 한국고용정보원의 자료와 노동연구원의 노동패널 자료이다. 한국고용정보원의 자료에는 고용보험 가입자만이 재직근로자 직업훈련의 대상이기 때문에 본 연구와 직접적인 비교는 어렵다(2008년 42.7%의 훈련참여율). 또한 노동연구원의 한국노동패널의 경우 '지난 조사 이후 현재의 교육훈련까지 포함하여 몇 개의 교육훈련을 받았거나 받고 계십니까?'라는 질문에 대하여 1998년부터 2007년 사이에 약 11%가 훈련을 받았다고 답하고 있어 본고의 경활 자료와 많은 차이를 보이고 있다(유경준 · 이철인[2009] 참조). 그러나 현재 이러한 차이의 정확한 이유를 판단하기는 용이하지 않다.

료 간에 큰 차이를 보이지는 않는다. 마찬가지로, 평균 나이도 약 42.0~43.1세, 그리고 평균 교육연수는 약 12.7~12.8년 정도로 각 자료 간에 큰 차이가 나타나지 않는다. 근로자의 근로시간, 직종, 업종, 사업장 규모 등 직장 관련 속성에 있어서도 표본 기간 내에서 매해 유사한 분포를 보이고 있다.

Ⅳ. 교육훈련의 효과

본 장에서는 앞의 Ⅱ장에서 설명된 통계분석 방법인 고정효과모형과 성향점수 매칭법의 추정 결과를 통하여 우리나라에서 직업훈련이 임금에 미치는 인과효과의 크기를 살펴본다.

1. 임금방정식 추정을 통한 효과 분석

먼저, <Table 3>에는 고정효과모형의 추정치가 보고되어 있다. Model 1은 2007년 3월과 2008년 3월을 연결한 '패널 1'과, 2008년 3월과 2009년 3월을 연결한 '패널 2'에 대하여 식 (1)을 추정한 결과이다. 반면에 Model 2는 2007년 3월, 2008년 3월, 2009년 3월의 세 경활조사를 연결한 '패널 3'에 대하여 식 (1)을 추정한 결과이다.[7] 식 (1)에서는 y_{it}가 '$\ln(w_{it}) - \ln(w_{it-1})$'로 정의되기 때문에, Model 1의 추정에서는 c_i가 독자적으로 식별되지 않으므로 모든 i에 대하여 '$c_i = 0$'이라고 가정하고 주어진 자료에 OLS를 적용한다.

'패널 1'을 이용한 모형 'Model 1-1'에 대한 추정치는 <Table 3>의 (1)열에 보고되어 있고, '패널 2'를 이용한 모형 'Model 1-2'에 대한 추정치는 표의 (2)열에 보고되어 있다. '패널 1'과 '패널 2'를 수직결합한(pooling) 자료를 이용한 모형 'Model 1-3'에 대한 추정치는 (3)열에 제시되어 있다. 마지막으로, (4)열과 (5)열에는 3년간의 결합 자료 '패널 3'을 이용하여 식 (1)을 추정한 추정 결과가 보고되어 있다. (4)열에는 '패널 3'에 대하여 '$c_i = 0$'을 가정한 후 OLS법을 적용했을 경우의 추정 결과가 제시되어 있고, (5)열에는 '$c_i = 0$'를 가정하지 않고 wihtin transformation을 적용한 추정 결과가 보고되어 있다.

식 (1)에서는 y_{it}가 '$\ln(w_{it}) - \ln(w_{it-1})$'의 차분으로 정의되기 때문에 <Table 3>에 보고된 γ의 추정치들은 종속변수로서 '$\ln(w_{it})$'를 이용하는 분석에 비하여 편의(bias)가 상대적으로 적을 가능성이 높다. 통계학적인 관점에서 볼 때, (5)열에 보고된 고정효과모형의 추정치가

7) 본고에서 계산된 모든 임금 관련 통계는 물가조정(2006=100)된 실질임금을 바탕으로 계산한 것이다.

〈Table 3〉 Fixed Effects Estimation of the Earnings Equation: A Difference in Logarized Monthly Earnings as the Dependent Variable

		model 1(0703~0903)			model 2	
		model 1-1 (0703~0803)	model 1-2 (0803~0903)	model 1-3 (0703~0903)	0703-0903, including separated workers (0703~0903)	
Estimation Method		ols	ols	ols	ols	fixed effect
		(1)	(2)	(3)	(4)	(5)
train	job training, trained=1	0.0466	0.0263	0.0365	0.0346	0.0468
		(0.0080)***	(0.0071)***	(0.0053)***	(0.0084)***	(0.0170)***
sex	gender, male=1	-0.0168	-0.0085	-0.0134	-0.0089	-
		(0.0090)*	(0.0085)	(0.0062)**	(0.0098)	
rel	head of the household, head=1	0.0000	-0.0162	-0.0082	-0.0051	-0.0969
		(0.0089)	(0.0083)*	(0.0061)	(0.0100)	(0.0560)*
marr	marital status, married=1	-0.0006	-0.0036	-0.0020	0.0111	0.0048
		(0.0088)	(0.0081)	(0.0060)	(0.0098)	(0.0951)
age	age	-0.0001	-0.0044	-0.0023	-0.0077	-0.0519
		(0.0022)	(0.0020)**	(0.0015)	(0.0025)***	(0.0686)
age2	age^2	-0.0000	0.0000	0.0000	0.0001	0.0004
		(0.0000)	(0.0000)*	(0.0000)	(0.0000)***	(0.0004)
eduyear	years of education	0.0002	0.0025	0.0013	0.0021	0.0207
		(0.0015)	(0.0014)*	(0.0010)	(0.0016)	(0.0149)
contyr	years of work (in a same job)	-0.0029	-0.0032	-0.0031	-0.0030	0.0239
		(0.0015)*	(0.0014)**	(0.0010)***	(0.0016)*	(0.0077)***
contyr2	year of work^2	0.0001	0.0001	0.0001	0.0001	-0.0006
		(0.0001)**	(0.0000)*	(0.0000)***	(0.0001)*	(0.0003)**
city	city dwelling, city dweller=1	0.0160	-0.0143	0.0013	0.0003	-
		(0.0089)*	(0.0085)*	(0.0061)	(0.0098)	
establishment size, occupation, industry		controlled	controlled	controlled	controlled	controlled
normal	Non-regular work, non-regular=1	-0.0145	-0.0384	-0.0266	-0.0336	-0.1235
		(0.0082)*	(0.0079)***	(0.0057)***	(0.0093)***	(0.0233)***
chg1	separation, separated=1	0.0253	-0.0065	0.0086	0.0119	0.0381
		(0.0089)***	(0.0079)	(0.0059)	(0.0098)	(0.0175)**
worktime	hours worked(month)	0.0009	0.0008	0.0009	0.0008	0.0028
		(0.0001)***	(0.0001)***	(0.0001)***	(0.0001)***	(0.0002)***
dum0803	dummy for March 2008, 2008=1	-	-	0.0100	0.0183	0.0056
				(0.0045)**	(0.0071)**	(0.0596)
Constant		-0.1004	-0.0413	-0.0748	0.0345	0.4679
		(0.0506)**	(0.0469)	(0.0346)**	(0.0589)	(2.7032)
Observations		10790	11401	22191	7924	7924
Number of id						3962
R-squared(fixed effect: within)		0.03	0.03	0.03	0.02	0.08

Note: Model 1 uses all samples with change in wage (% change), while model 2 only uses samples that can be set up as a panel. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

가장 편의(bias)가 적을 것으로 예상된다. 그러나 우려되는 한 가지 문제는 고정효과모형의 추정에서 사용되는 관측치의 수가 적어서 소규모 샘플 편의가 발생할 수 있다는 점이다. 이 점을 고려하여 본 논문에서는 상이한 방식으로 추정된 '교육훈련'의 임금효과를 동일한 비중으로 해석한다. 이 추정치들이 모형 가정상의 차이에도 불구하고 꽤 좁은 범위 안에 모여 있는 경우, 우리는 추정치들이 신뢰할 만하다고 판단할 수 있을 것이다.

먼저, 'Model 1-1'과 'Model 1-2'의 결과에 따르면, 지난 1년 동안에 이수한 교육훈련으로 인하여 올해의 월평균 임금수준은 평균 2.63~4.66% 정도 상승한다.[8] '패널 1'과 '패널 2'를 수직결합한 자료를 이용하는 'Model 1-3'의 추정 결과에서도 비슷한 결과가 나타나서 교육훈련은 개인의 월평균 임금수준을 약 3.65% 정도 상승시킨다. 이 정도 크기의 임금 상승효과는 3년간의 결합 자료를 이용하는 'Model 2'에서도 비슷하게 관측된다. 'Model 2'의 자료에서 발견되는 '교육훈련'에 대한 OLS의 추정치는 평균 3.46% 정도의 임금 상승효과를, '확장된' 고정효과모형의 추정치는 평균 4.68% 정도의 임금 상승효과를 보여준다. 종합하면, 지난 1년 동안에 이수한 교육훈련으로 인하여 올해의 월평균 임금수준은 약 2.6~4.7% 정도 상승하는 것으로 추정된다. 그리고 이 통계치들은 최소 5%의 유의수준에서 통계적으로 유의하다.[9]

2. 매칭법을 통한 직업훈련의 효과 분석

가. 교육훈련 참가 여부

<Table 4>에는 매칭에 사용될 성향점수(propensity score)를 추정하는 프로빗 모형의 추정치들이 보고되어 있다. 한 개인이 t기에 교육훈련에 참여할지의 여부는 (t-1)기의 특성에 영향을 받는다는 전제하에 성향점수는 다음의 식을 통하여

8) 본고의 고정효과모형에서는 종속변수가 로그임금의 차분으로 설정되어 있다. 하지만 훈련 여부에 대한 추정계수는 수준변수를 종속변수로 사용하는 통상적인 경우와 동일하게 훈련 여부에 따른 임금 상승의 퍼센티지로 해석될 수 있다. 왜냐하면, 본고의 실증모형은 전기의 로그임금 수준을 좌변에서 우변으로 옮기고 그 계수를 1로 가정한 모형으로 변환될 수 있기 때문이다. 차분 대신에 전기의 임금을 우변의 설명변수로 사용하는 모형을 생각해 볼 수 있다. 그러나 이 모형은 기본적으로 동태패널모형(dynamic panel data model)으로서 추정방법이 불필요하게 복잡해질 가능성이 있어서, 이 문제를 회피하기 위하여 본고에서는 차분을 종속변수로 사용하는 모형을 채택하였다.

9) <Table 3>의 (4)열과 (5)열의 결과로부터 within transformation의 훈련효과 추정치(0.0468)가 OLS의 추정치(0.0346)보다 크다는 점을 확인할 수 있다. 두 추정치의 크기 차이로부터, 전기 대비 임금 상승률이 낮을 가능성이 있는 근로자들이 주로 교육훈련을 받는다는 점을 유추해 볼 수 있다. 그러나 두 추정치 모두가 상대편 추정치의 95% 신뢰구간 안에 포함되기 때문에 위의 예상이 통계적으로 확정적이라고 판단하기는 어렵다.

추정되었다.

$$\Pr(Z_t = 1 \mid X_{t-1}) = \Phi[h(X_{t-1})] \qquad (5)$$

Matching 1은 '패널 1'에 대하여, Matching 2는 '패널 2'에 대하여 식 (5)를 추정한 결과이다. Matching 3은 '패널 1'과 '패널 2'를 수직결합한 자료에 대한 식 (5)의 추정 결과이다.

<Table 4>의 결과에 따르면, 교육훈련의 이수 가능성은 전기의 임금수준이 높을수록, 여성, 가구주, 기혼자일수록, 그리고 학력이 높을수록, 나이가 적을수록, 근속연수가 길수록 높다. 그러나 성별 및 가구주 여부의 계수 추정치는 항상 유의하지는 않다. 2차항을 추가하는 경우, 연령의 제곱항은 유의하지 않아서 나이의 효과는 선형으로 판단되고, 근속연수의 제곱항은 유의한 음의 값을 가지므로 교육훈련의 확률은 근속연수에 따라 체감적으로 증가한다.

<Table 4>에 명시적으로 보고되지는 않았지만, 전반적으로 사업체 규모가 큰 사업장 근로자일수록 교육훈련 확률이 높으나, 5인 미만 사업장에 비해 5~29인 사업장에서는 오히려 교육훈련 확률이 낮은 것으로 나타난다. 직종의 경우 서비스직은 기준집단(준전문직과 사무직)에 비해 교육훈련을 경험할 확률이 유의하게 높으며, 농림어업, 기능, 조립 등 기타 직종과 단순노무의 경우 교육훈련 확률이 유의하게 낮다. 그리고 관리·전문 직종의 계수 추정치는 유의하게 나타나지 않는다.

업종의 경우에는 제조업과 비교하여 금융보험·사업서비스업, 공공행정·교육서비스·보건업에서 교육훈련을 경험할 확률이 유의하게 높다. 나머지 업종은 제조업에 비해 상대적으로 교육훈련을 경험할 확률이 낮은 것으로 나타났으나, 도소매·숙박을 제외한 나머지 업종의 경우 계수 추정치가 일관되지 않으며 유의하지 않다.

나. 임금효과

<Table 5>에는 성향점수 매칭법을 통해 추정된 교육훈련의 임금효과가 제시되어 있다.[10][11] 2007년 3월과 2008년 3월을

10) 성향점수(Propensity Score)를 이용한 매칭 시 STATA의 pscore, attnd(Sascha O. Becker and Andrea Ichino) 등을 이용하였다. ATT의 표준오차와 t-value는 다음의 계산식을 이용하여 계산하였고, stratification 방법 이용 시 bootstrapped standard error를 사용하였다.

$$\text{t-value} = \frac{(\overline{X} - \overline{Y})}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},\ t(n_1 + n_2 - 2),\ s_p^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

11) <Appendix Table 2>에서는 가정 2의 balancing property를 검정하기 위하여 매칭 전과 매칭 후 처치집단과 비교집단 간 근로자 특성의 차이를 계산하였다. 매칭 후 처치집단과 비교집단 간에 근로자 특성의 유의한 차이가 발견되지 않으므로 가정 2가 대체로 성립한다고 판단된다.

〈Table 4〉 Determinants of Work-related Training

	Matching 1 (0703~0803)		Matching 2 (0803~0903)		Matching 3 (0703~0903)	
Explanation for Variables	Variable Average (Standard Deviation)	probit	Variable Average (Standard Deviation)	probit	Variable Average (Standard Deviation)	probit
log(wage) from previous period	0.03	0.4275	0.02	0.4471	0.02	0.4287
	(0.34)	(0.0356)***	(0.33)	(0.0342)***	(0.32)	(0.0245)***
gender, male=1	0.57	-0.0469	0.57	-0.0618	0.58	-0.0609
	(0.49)	(0.0413)	(0.50)	(0.0384)	(0.49)	(0.0279)**
head of household, head=1	0.60	-0.0005	0.60	0.0114	0.60	0.0026
	(0.49)	(0.0400)	(0.49)	(0.0374)	(0.49)	(0.0273)
marital status, married=1	0.71	0.0884	0.70	0.0324	0.71	0.0603
	(0.45)	(0.0400)**	(0.46)	(0.0363)**	(0.45)	(0.0268)**
age	12.75	-0.0160	12.68	-0.0131	12.82	-0.0171
	(3.31)	(0.0104)	(3.31)	(0.0095)	(3.31)	(0.0070)**
age^2	41.13	0.00002	40.96	0.00002	41.30	0.00005
	(11.31)	(0.0001)	(11.39)	(0.0001)	(11.23)	(0.0001)
length of education	1819.89	0.0332	1807.54	0.0309	1831.58	0.0327
	(984.50)	(0.0073)***	(987.83)	(0.0064)***	(981.25)	(0.0048)***
year of work	5.89	0.0547	5.81	0.0595	5.96	0.0593
(in the same job)	(7.40)	(0.0061)***	(7.35)	(0.0057)***	(7.44)	(0.0041)***
years of work^2	89.40	-0.0011	87.82	-0.0011	90.89	-0.0011
	(188.76)	(0.0002)***	(185.27)	(0.0002)***	(192.00)	(0.0001)***
has lived in a city	0.83	-0.0152	0.83	-0.2314	0.84	-0.1390
in the previous period	(0.37)	(0.0393)	(0.38)	(0.0366)***	(0.37)	(0.0266)***
non-regular job	0.33	-0.0664	0.33	-0.1685	0.32	-0.1118
	(0.47)	(0.0361)*	(0.47)	(0.0343)***	(0.47)	(0.0247)***
hours worked (months)	197.19	-0.0017	197.19	-0.0020	196.26	-0.0018
	(48.66)	(0.0004)**	(48.66)	(0.0003)***	(48.26)	(0.0002)***
dummy for March 2008,	-		-		0.49	-0.2443
2008=1					(0.50)	(0.0208)***
constant	-	-2.7031	-	-2.4123	-	-2.3077
		(0.2598)***		(0.2350)***		(0.1725)***
establishment size, occupation, industry		controlled		controlled		controlled
Observations	10,790		11,401		22,191	
Log-likelihood	-5260.2		-6072.1		-11398.5	

Note: For every explanatory variable we used the value from the year before. This includes the workers who left the job, Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1.

〈Table 5〉 Matching Estimates of Average Treatment Effects on the Treated (ATT): A Difference in Logarized Monthly Earnings as the Outcome Variable

(Unit: person)

	matching method	treatment group	control group	ATT	standard deviation	t-value
0703~0803	stratification	2,956	7,827	0.085	0.012	7.10***
	nearest	2,956	1,874	0.076	0.020	3.78***
	radius matching	2,024	3,610	0.260	0.017	15.22***
	kernel matching	2,956	7,828	0.111	-	-
0803~0903	stratification	3,822	7,572	0.075	0.008	8.99***
	nearest	3,822	2,187	0.058	0.019	2.98***
	radius matching	2,596	3,933	0.182	0.016	11.51***
	kernel matching	3,822	7,571	0.090	-	-
0703~0903	stratification	6,778	15,396	0.075	0.007	10.43***
	nearest	6,778	4,061	0.075	0.014	5.42***
	radius matching	5,680	11,038	0.287	0.010	27.89***
	kernel matching	6,778	15,397	0.098	-	-

Note: *** p<0.01, ** p<0.05, * p<0.1.

연결한 '패널 1'에 대한 stratification, nearest neighbor와 kernel matching의 결과에 의하면, 교육훈련의 영향으로 개인의 월평균 임금은 평균 약 7.6~11.1% 정도 상승한다. 그러나 radius matching에 의한 추정치는 이보다 훨씬 큰 26.0% 정도의 수준으로 상당히 높게 나타난다. 2008년 3월과 2009년 3월을 연결한 '패널 2'를 이용하는 경우에 stratification, nearest neighbor와 kernel matching의 결과는 교육훈련의 월평균 임금 상승효과가 평균 5.8~9.0% 정도 수준임을 보여준다. '패널 1'의 경우와 마찬가지로 '패널 2'에서도 radius matching의 추정치는 상당히 큰 임금효과(18.2%)를 보여준다. 마지막으로, '패널 3'에 대한 stratification, nearest neighbor와 kernel matching의 결과는 교육훈련의 임금효과가 평균 7.5~9.8% 정도임을 보여준다. Radius matching에 의한 임금효과는 약 28.7% 정도로서 상당히 크게 나타난다.[12)]

본 절에서 stratification, nearest neighbor와 kernel matching법을 통해 추정된 교육훈련의 임금 상승효과 7.5~9.8%('패널 3'

12) Radius matching의 결과는, 그 길이와 관련없이 또한 common support의 영역 내의 추정에도 불구하고,

기준)는 앞 절에서 OLS 및 고정효과모형을 통해 추정된 교육훈련의 임금 상승 효과(2.6~4.7%)에 비해 약 2배 정도 높다. 두 추정방법이 서로 상이한 크기의 임금효과를 암시하고 있어서 우리나라 노동시장에서 발생하는 교육훈련의 정확한 임금효과를 제시하는 데에는 다소간의 어려움이 있다. 그럼에도 불구하고, 고정효과모형과 매칭법의 결과들을 종합하면, 지난 1년간 이수한 교육훈련은 개인의 월평균 임금을 평균 약 2.6~9.8% 정도 상승시키는 것으로 추정된다. 그리고 이 추정치들은 최소 5%의 유의수준에서 모두 유의하다.

V. 결 론

본고는 임금근로자에 대한 직업훈련의 임금효과를 추정하기 위하여 「경제활동인구조사」 자료에 대하여 고정효과모형과 매칭기법을 적용하였다. 분석 결과, 고정효과모형의 경우 2008년에는 4.7%, 2009년에는 2.6%의 유의한 임금 상승효과가 추정되었고, 2개년을 평균하면 직업훈련은 약 3.7%의 임금 상승효과가 있는 것으로 추정되었다. 한편, 매칭을 이용한 추정 결과는 이보다 대체로 큰 약 7.5~9.8%의 임금 상승효과를 보여준다. 종합하면, 우리나라에서 직업훈련은 근로자의 월평균 임금을 평균 2.6~9.8% 정도 상승시키는 것으로 판단된다.

그러나 외국문헌에서의 훈련과 임금상승 간의 관계는 우리나라에 대한 본고의 발견과 일치하지 않는 경우도 많이 존재한다. 외국의 경우 기업이 근로자의 훈련비용을 부담하는 것을 흔히 발견할 수 있다.[13] 노동시장이 완전경쟁적이라면 Becker의 이론에 따라 훈련비용을 근로자가 모두 부담하여야 한다. 따라서 기업이 훈련비용을 부담한다는 것은 근로자들이 낮은 임금을 통해 훈련비용의 일부를 부담한다는 것을 포함하는 의미가 되어 노동수요 독점적인 요소가 있는 불완전 노동시장이 설정되어 있음을 의미한다. 즉, 기업이 훈련비용을 부담하는 경우 기업은 훈련비용의 회수를 염두에 두기 때문에 그들이 노동수요를 어느 정도 독점하여 훈련으로 인한 생산성의 상승보다 임

기존의 문헌과 본 논문의 다른 결과와 대비해서 부자연스럽게 큰 임금효과를 보여준다. 이론적으로 radius matching의 결과가 배제될 타당한 이유는 존재하지 않지만, 다른 분석 결과들과 너무 이질적이라는 의미에서 매칭법의 실증 결과를 요약하는 데 있어서는 radius matching의 결과를 제외하기로 한다. Radius matching의 결과가 다른 매칭의 결과와 다른 이유를 여러 방면으로 찾으려고 노력하였으나, 아직까지 그 타당한 이유를 발견하지 못하였다.

13) 이는 우리나라의 경우도 누가 훈련비용을 부담하느냐에 따라 훈련의 효과가 상이할 수 있음을 의미한다. 그러나 본고의 분석에서 사용된 자료는 훈련비용의 부담주체를 식별하지 못하는 한계를 가지고 있다. 훈련비용 부담주체에 따른 훈련효과의 차이를 분석한 논문은 채창균(2009)을 참조할 수 있다.

금 상승이 낮아야만 근로자에 대한 훈련 투자가 가능하다. 따라서 근로자들이 낮은 임금을 감수하면서 훈련에 투자한다면, 임금 프리미엄과 훈련량 사이에 양의 상관관계가 과연 존재하는지 살펴보아야 한다. 외국의 많은 연구들은 이러한 주장을 완전히 뒷받침하고 있지는 않다(Brunello and Paola[2004]).

또한 Bassanini and Brunello(2003)는 유럽의 가구패널(European Community Household Panel) 자료를 이용하여, 국가별로 훈련받은 근로자들과 훈련 받지 않은 근로자들 사이의 중위임금(median wage) 상승의 차이를 이용하여 훈련에 대한 임금 프리미엄의 크기를 측정하였다. 이를 통해 그들은 임금압축현상(wage compression)으로 인하여, 훈련을 받은 근로자가 훈련을 받지 않은 근로자보다 유의하게 높은 임금을 받지는 않는다고 보고하였다.

또한 선택편의를 제거한 계량기법을 사용한 연구들은 훈련으로 인한 임금 상승 효과가 양이기는 하지만 그리 크지 않다는 것을 보여주고 있다(Brunello and Paola[2004]). 뿐만 아니라 네덜란드에서 훈련을 받고 싶었으나 우연히도(random event) 받지 못한 근로자를 비교 대상으로 하여 일과 관련된 훈련의 보상을 측정한 결과, 훈련을 받은 근로자와 다르지 않음을 보인 논문도 존재한다(Leuven and Oosterbeek[2008]).

종합하면, 훈련과 임금 상승 간의 관계는 국가별로 훈련에 대한 정의 문제와 훈련의 정도가 다르기 때문에 상이한 결과를 보이고 있다. 훈련을 적게 하는 나라에서 훈련의 수익률이 높지만, 이 관계는 훈련변수의 내생성이 제대로 통제되지 않았기 때문이며, 훈련의 내생성이 잘 통제될수록 훈련의 임금 상승효과가 줄어들거나 유의하지 않은 결과를 보여주는 국가들도 있다(Brunello, Garibaldi, and Wasmer[2007]).

필자들은 이러한 점들을 충분히 고려하여 향후 우리나라에서도 노동시장 구조와 훈련 및 임금 상승 간의 관계에 대한 폭넓고 심층적인 연구가 계속 이루어지기를 기대한다.

참 고 문 헌

(in Korean)

유경준・이철인, 「직업훈련의 임금효과 분석: 노동패널 자료를 중심으로」, mimeo, 한국개발연구원, 2009.

이상호, 「교육훈련 기회와 노동시장 성과」, 『노동리뷰』, 2005년 8월호, pp.61~74.

채창균, 「비정규직의 직업훈련 참여 실태와 성과」, 유경준 편, 『비정규직 문제 종합연구』, 제6장, 연구보고서 2009-13, 한국개발연구원, 2009.

Kang, Soon Hie and Heung Sung Nho, "The Effect of Job Training in Korea on Employment and Wage," *The Korean Labor Economic Association Dissertation*, Vol. 23, No. 2, 2000, pp.127~151.

Kim, Ahn Kook, "A Study on the Individual Wage Effect of Training," *The Korean Labor Economic Association Dissertation*, Vol. 25, No. 1, 2002, pp.131~160.

(in English)

Bassanini, A., A. L. Booth, G. Brunello, M. De Paola, and E. Leuven, "Workplace Training in Europe," IZA (Institute for the Study of Labor) Discussion Papers, No. 1640, 2005.

Basasnini, A. and G. Brunello G., "Is Training More Frequent When Wage Compression Is Higher? Evidence from the European Community Household Panal," OECD, IZA Discussion Papers, No. 839, 2003.

Blundell, R., L. Dearden, C. Meghir, and B. Sianesi, "Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy," *Fiscal Studies*, Vol. 20, No. 1, 1999, pp.1~23.

Brunello, M., P. Garibaldi, and E. Wasmer, *Education and Training in Europe*, Oxford University Press, 2007.

Brunello, Giorgoi and Maria De Paola, "Market Failures and The Under-provision of Training," Papered prepared for the joint EC-OECD Seminar on Human Capital and Labor Market Performance, held in Brussels on Dec 8, 2004.

Card, D., "The Causal Effect of Education on Earnings" in Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Amsterdam: North Holland, 2000.

Heckman, J. J., L. Lochner, and C. Taber, "Tax Policy and Human-Capital Formation," *American Economic Review*, Vol. 88, No. 2, 1998, pp.293~297.

Leuven, E. and H. Oosterbeek, "An Alternative Approach to Estimate the Wage Returns to

Private-sector Training," *Journal of Applied Econometrics*, Vol. 23, No. 4, 2008, pp.423~434.
Rosenbaum, Paul and Donald Rubin, "The Central Role of Propensity Score in Observational Studies for Causal Effects," *Biometrika*, Vol. 70, No. 1, 1983, pp.41~55.
Smith, J. A. and P. E. Todd, "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, Vol. 125, No. 1~2, 2005, pp.305~353.

〈Appendix Table 1〉 Fixed Effects Estimation of the Earnings Equation: A Difference in Hourly Wages as the Dependent Variable

	Model 1				
	Model 1-1 (0703~0803)	Model 1-2 (0803~0903)	Model 1-3 (0703~0903)	Model 2 (0703~0903)	
estimation method:	ols	ols	ols	ols	fixed effect
	(1)	(2)	(3)	(4)	(5)
job training, trained=1	0.0490 (0.0078)***	0.0254 (0.0074)***	0.0368 (0.0054)***	0.0322 (0.0085)***	0.0513 (0.0180)***
gender, male=1	0.0063 (0.0088)	-0.0025 (0.0088)	0.0010 (0.0062)	0.0069 (0.0099)	-
head of household, head=1	0.0061 (0.0087)	-0.0017 (0.0086)	0.0018 (0.0061)	0.0040 (0.0101)	-0.0580 (0.0593)
marital status, married=1	-0.0150 (0.0086)*	-0.0139 (0.0084)*	-0.0146 (0.0060)**	-0.0092 (0.0099)	0.0153 (0.1008)
age	-0.0035 (0.0021)*	0.0014 (0.0021)	-0.0010 (0.0015)	-0.0039 (0.0025)	-0.0853 (0.0725)
age^2	0.0000 (0.0000)	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0004)
years of education	-0.0008 (0.0015)	0.0015 (0.0014)	0.0002 (0.0010)	0.0014 (0.0016)	0.0040 (0.0158)
years of work	-0.0005 (0.0015)	-0.0025 (0.0015)*	-0.0015 (0.0011)	-0.0018 (0.0016)	0.0292 (0.0082)***
years of work^2	0.0000 (0.0001)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0001)	-0.0006 (0.0003)**
non-regular work, non-regular=1	0.0031 (0.0078)	-0.0164 (0.0079)**	-0.0068 (0.0056)	-0.0086 (0.0092)	-0.0026 (0.0244)
city dwelling, city dweller=1	0.0083 (0.0087)	-0.0119 (0.0088)	-0.0017 (0.0062)	-0.0004 (0.0100)	-
separation, separated=1	0.0053 (0.0087)	-0.0294 (0.0082)***	-0.0127 (0.0060)**	-0.0065 (0.0099)	0.0230 (0.0186)
dummy for March 2008, 2008=1	-	-	0.0101 (0.0045)**	0.0267 (0.0073)***	-0.0522 (0.0631)
constant	0.1313 (0.0476)***	0.0129 (0.0472)	0.0676 (0.0337)**	0.1078 (0.0574)*	3.4656 (2.8624)
establishment size, occupation, industry	controlled	controlled	controlled	controlled	controlled
Observations	10,790	11,401	22,191	7,924	7,924
Number of id	-	-	-	-	3,962
R-squared	0.01	0.02	0.01	0.02	0.04

Note: Model 1 uses all samples with change in wage (% change), while model 2 only uses samples that can be set up as a panel. Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1.

<Appendix Table 2>에서는 매칭 전후의 근로자 특성의 비교치를 제시한다. 이는 비모수적 방법인 매칭 분석에서 가장 중요시하는 비교집단과 처치집단의 임의 추출 여부, 즉 어느 정도 비교그룹과 처치그룹이 유사해졌는지를 제시하기 위한 것이다. 매칭 분석을 수행하기 위해서는 매칭과정이 끝난 후에 두 그룹의 특성들의 차이가 통계적으로 유의하지 않아야 한다. 여기서는 대표적으로 0703~0903 자료를 이용하여 Nearest Neighbors Matching 방법을 이용한 경우를 제시한다. 매칭 후의 결과를 보면, 일부 더미변수를 제외하고 모든 변수들에서 통계적으로 유의한 차이가 나타나고 있지 않다. 이는 본 매칭 수행이 성공적으로 이루어졌다고 볼 수 있는 근거를 제공한다고 할 수 있다.

〈Appendix Table 2〉 Comparisons of Worker Characteristics before and after Matching

– Means of Worker Characteristics before Matching

Variable		0703~0903 data		
		Untrained N=15413 (Mean)	Trained N=6778 (Mean)	t-value[1)]
Gender		0.548	0.634	-11.93***
Head of Household		0.584	0.642	-8.17***
Marital Status		0.687	0.759	-10.79***
Years of Education		12.20	14.01	-38.78***
Age		41.78	39.67	12.85***
Age^2		1886.86	1667.60	15.36***
Years of Work		4.384	9.305	-47.94***
Years of Work^2		57.77	161.32	-38.90***
City Dweller		0.832	0.838	-1.08
Size	5~ 29	0.414	0.286	18.26***
	30~ 99	0.186	0.280	-15.77***
	100~299	0.086	0.152	-14.76***
	More than or equal to 300	0.028	0.054	-9.78***
Occupation	Management · Professional	0.141	0.267	-22.86***
	Service · Sales	0.173	0.138	6.54***
	Others(agriculture and forestry · fishing, technic etc.)	0.255	0.176	12.86***
	Simple Labor	0.197	0.066	25.10***
Industry	Agriculture and Forestry · Mining	0.017	0.003	8.20***
	Electricity · Gas · Water · Construction	0.063	0.045	5.43***
	Wholesale and Retails · Accommodation	0.193	0.067	24.19***
	Transportation · Communication	0.111	0.101	2.16**
	Financial insurance · Business	0.160	0.194	-6.11***
	Administration · Education · Health	0.100	0.201	-20.77***
Others excluding manufacturing		0.076	0.045	8.63***
Non-regular Work		0.365	0.235	19.25***
Hours Worked		198.94	190.17	12.51***
Availability of '08 data		0.508	0.436	9.93***

Note: difference = mean(train=0)-mena(train=1), H_0: difference=0

– Means of Worker Characteristics after Matching

Variable		0703~0903(By using Nearest Neighbors Matching)		
		Untrained N=6778 [2] (Mean)	Trained N=6778 (Mean)	t-value
Gender		0.640	0.634	0.84
Head of Household		0.650	0.642	0.90
Marital Status		0.769	0.759	1.46
Years of Education		14.02	14.01	0.24
Age		39.93	39.67	1.54
Age^2		1691.18	1667.60	1.70*
Years of Work		9.34	9.31	0.20
Years of Work^2		162.47	161.32	0.28
City Dweller		0.836	0.829	-0.33
size	5~ 29	0.286	0.286	-0.08
	30~ 99	0.286	0.280	0.71
	100~299	0.151	0.152	-0.22
	More than or equal to 300	0.059	0.054	1.26
Occupation	Management · Professional	0.274	0.267	0.87
	Service · Sales	0.131	0.138	-1.21
	Others(agriculture and forestry · fishing, technic etc.)	0.176	0.176	-0.02
	Simple Labor	0.065	0.066	-0.07
Industry	Agriculture and Forestry · Mining	0.003	0.003	-0.15
	Electricity · Gas · Water · Construction	0.042	0.045	-0.63
	Wholesale and Retails · Accommodation	0.063	0.067	-1.08
	Transportation · Communication	0.105	0.101	0.73
	Financial insurance · Business	0.206	0.194	1.72*
	Administration · Education · Health	0.203	0.201	0.26
Others excluding manufacturing		0.049	0.045	1.14
Non-regular Work		0.242	0.235	0.99
Hours Worked		190.08	190.17	-0.12
Availability of '08 data		0.453	0.436	2.01**

Note: When implementing the Nearest Neighbors Matching, we can either or not allow overlaps. When overlaps are allowed, one unit of control group will be matched to multiple units of treatment groups. Therefore, we need to consider the overlapped weights in order to match the number of units in the treatment group and the control group. This table also took weights into account when mean difference test was carried out.