

## Supplemental Materials for “Causal Effect of Urban Parks on Children’s Happiness”

Nayeon Kwon<sup>a</sup> Chanmin Kim<sup>1,a</sup>

<sup>a</sup>Department of Statistics, SungKyunKwan University

### Appendix A: 반응변수 히스토그램

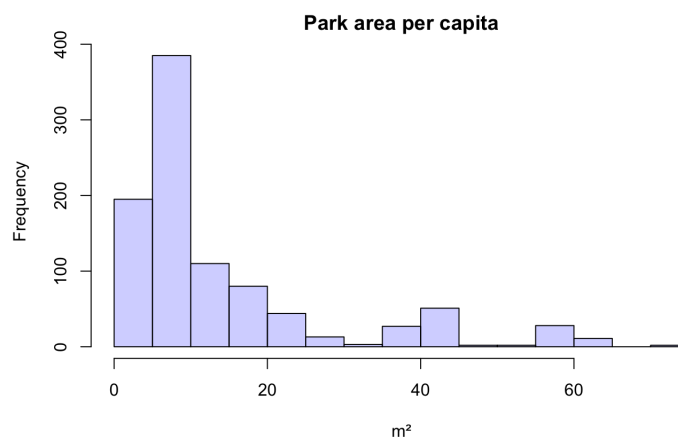


Figure A.1: Histogram of park area per capita

### Appendix B: 유행비순환그래프 (Directed Acyclic Graph; DAG)

유행비순환그래프 (Directed Acyclic Graph; DAG)는 인과관계를 구조적으로 표현할 수 있는 도구이다. 특히, 교란변수의 집합을 찾을 때 유용하다. 유행비순환그래프는 그래프 내 변수들 간의 잠재적인 상관관계 도출 시 사용될 수 있다. 유행비순환그래프는 *D-separation*과 *backdoor criterion*을 바탕으로 도출한다. *D-separation*의 설명에 앞서 다음의 유행비순환그래프를 살펴보자. 1과 2는 chains, 3은 fork, 그리고 4는 collider 관계를 나타낸다:

1.  $Z \rightarrow X \rightarrow Y$
2.  $Z \leftarrow X \leftarrow Y$
3.  $Z \leftarrow X \rightarrow Y$
4.  $Z \rightarrow X \leftarrow Y$

This work is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2020R1F1A1A01048168, NRF-2022R1F1A1062904).

<sup>1</sup> Corresponding author: Assistant Professor, Department of Statistics, SungKyunKwan University, 25-2 Seonggyungwan-ro, Jongno-gu, Seoul 03063, Korea. E-mail: [chanmin.kim@skku.edu](mailto:chanmin.kim@skku.edu)

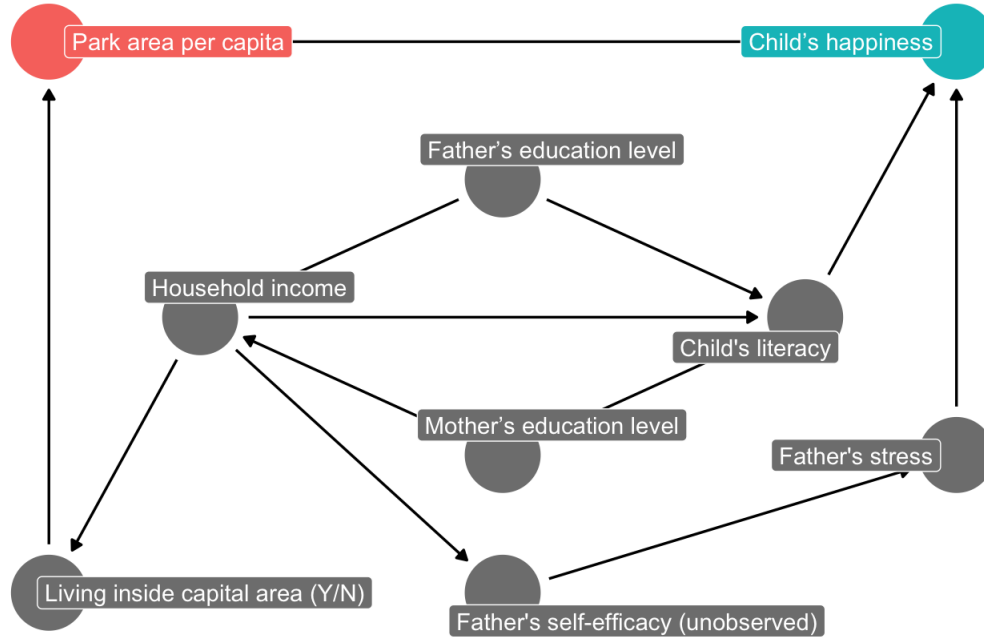


Figure B.1: DAG

*D*-separation은 주어진  $X$  집합에 대하여  $Z$ 와  $Y$ 의 조건부 독립을 식별하는 기준이다.  $Z$ 와  $Y$ 가 *D*-separate 되었다는 것은  $Z \perp Y \mid X$ 를 만족함을 뜻한다. Collider 4를 제외한 1, 2, 그리고 3의 경우  $Z \not\perp Y$ 이고  $Z \perp\!\!\!\perp Y \mid X$ 이다. 반면, 4는  $Z \perp\!\!\!\perp Y$ 이고  $Z \not\perp Y \mid X$ 이다 (Shalizi, 2013).

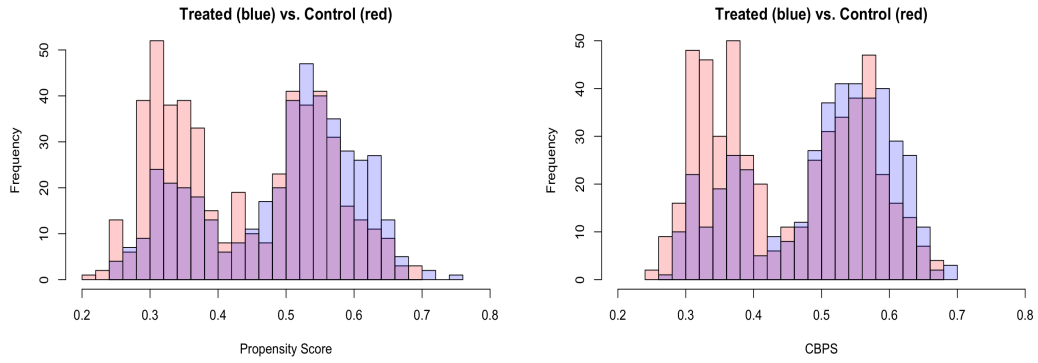
다음으로 backdoor criterion은 주어진  $X$  집합에 대하여  $Z$ 와  $Y$  간의 인과효과를 비모수적으로 식별하기 위한 기준이다.  $X$ 에 대한 backdoor criterion은 다음과 같다:

1.  $Z$ 의 descendent는  $X$  집합에 포함되지 않는다. 이는  $X$  집합이 직접적·간접적으로  $Z$ 에 의해 영향을 받지 않음을 의미한다.
2.  $X$ 는  $Z$ 로 들어오는  $Z$ 와  $Y$  간의 모든 경로를 *D*-separate한다.

만약 backdoor criterion를 만족하는 공변량들의 집합  $C$ 가 존재한다면 이를 통제함으로써 인과효과를 추정할 수 있다.

Figure B.1는 교란변수를 선택하기 위하여 도출한 유향비순환그래프이다. 처치변수 (park area per capita)와 반응변수 (child's happiness)를 기준으로 한 변수들 간의 관계는 다음과 같다. 거주지가 수도권에 위치하는 경우 (living inside capital area (Y/N)) 비교적 1인당 도시공원 면적이 좁은 경향이 있다. 그리고 수도권의 가구 소득 (household income)은 비수도권보다 높게 나타난다 (Kim 등, 2021). 또한 부모의 학력이 높을수록 가구 소득이 높아진다 (OECD, 2014). 어머니의 교육 수준에 따라 유아의 언어능력은 차이를 보이며 (Lim, 2019), 유아의 문해 및 언어능력이 높을수록 유아의 행복감은 높다 (Lee, 2017). 가구 소득은 관찰되지 않은 공변량인 아버지의 자기 효능감을 매개로 하여 양육 스트레스에 영향을 미치며 (Son과 Han, 2016), 아버지의 양육 스트레스가 낮을수록 유아의 행복감은 높게 나타난다 (Lee, 2017). 그러므로 backdoor criterion에 따라  $C$  집합에 포함하여 통제해야 하는 교란변수는 수도권 거주 여부 변수, 가구 월 평균 소득, 아버지 최종학력, 어머니 최종학력, 유아 문해 및 언어능력, 그리고 아버지 양육 스트레스 변수이다.

### Appendix C: 본 분석의 성향점수/CBPS 히스토그램



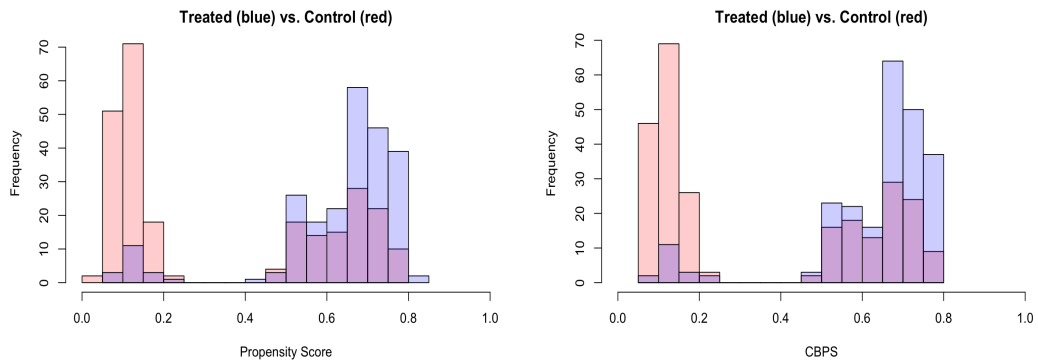
(a) Histograms of propensity score distribution

(b) Histograms of CBPS distribution

Figure C.1: Histograms of main analysis

### Appendix D: 민감도 분석 참고 자료

#### 성향점수/CBPS 히스토그램



(a) Histograms of propensity score distribution

(b) Histograms of CBPS distribution

Figure D.1: Histograms of sensitivity analysis

#### Covariate balance tables

## Appendix E: 추가 분석

세계보건기구 (World Health Organization; WHO)의 1인당 도시공원면적 권고기준인  $9\text{m}^2$  (An 등, 2013)를 처치변수 할당 기준으로 하여 추가 분석을 진행하였다. 표본의 크기는 처리군과 대조군 각각 420개, 533개, 합계 953개이다.

### 회귀모형을 이용한 조정 분석 결과

다음의 E.1은 회귀모형을 이용한 조정의 분석 결과를 나타낸다. 추정 결과 ATE는  $-0.015$ 이며, 표준오차는  $0.158$ 이다. 그리고 95% 신뢰 구간은  $[-0.325, 0.295]$ 이다. 95% 신뢰 구간의 상한과 하한 사이에 0을 포함하므로 유의미한 인과 효과가 존재하지 않는다. 이는 앞선 분석들과 동일한 결과이다.

### 가중치 기법 분석 결과

다음의 Figure E.1은 추가 분석의 그룹별 성향점수/CBPS 빈도 분포를 나타내는 히스토그램이다. 성향점수와 CBPS 모두 그룹 간의 오버랩 (overlap)이 잘 형성되어 있는 것을 확인할 수 있다.

다음의 Table E.2는 성향점수를 이용한 가중치 기법 분석 결과이다. 역확률 가중치 기법 (IPW), 안정화된 역확률 가중치 기법 (SIPW), 그리고 이중 강건 추정기법 (DR)의 점 추정치는 각각  $-0.152$ ,  $-0.038$ ,  $-0.039$ 로 일관적인 음수로 나타났다. 세 방법의 구간추정 결과 95% 신뢰구간의 상한과 하한 사이에 0을 포함한다. 그러므로 도시공원의 면적과 유아 행복감 간의 유의미한 인과효과는 존재하지 않는다.

Table E.3를 통해 CBPS를 이용한 가중치 기법 분석 결과를 확인할 수 있다. 위의 결과와 동일하게 세 방법을 이용하여 추정한 ATE는 일관적으로 음수였다. 그러나 역확률 가중치 기법 (IPW)의 경우 표준오차 ( $0.657$ )가 비교적 크게 나타났다. 모든 95% 신뢰구간 사이에 0을 포함하므로 두 변수 간 유의미한 인과효과가 존재한다고 할 수 없다.

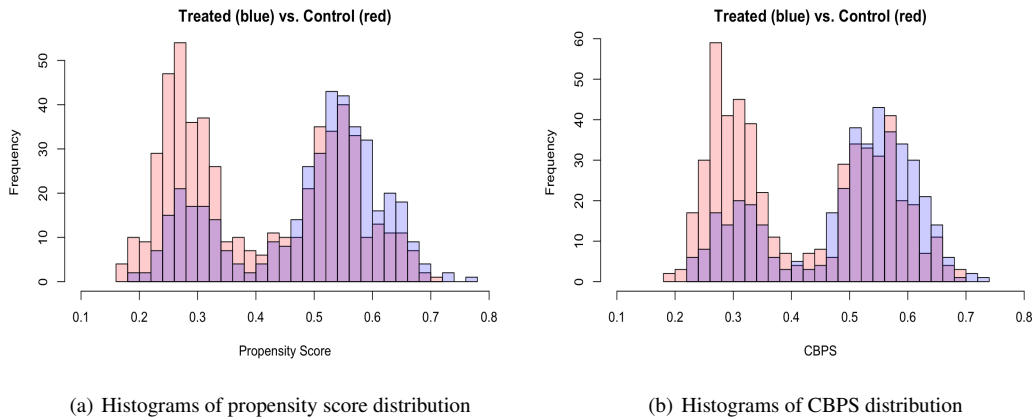


Figure E.1: Histograms of additional analysis

### 매칭 분석 결과

다음의 Figure E.2는 12개의 공변량에 대한 매칭 전/후의 표준화된 평균 차의 절댓값을 보여준다. 매칭 전 처리군과 대조군의 표본 크기는 각각 420개, 533개이다. 표준화된 평균 차의 절댓값을 기준으로 하였을 때, 수도권

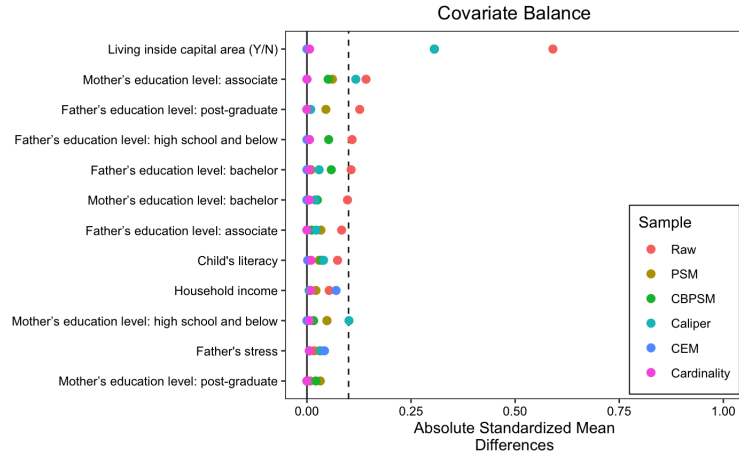


Figure E.2: Absolute SMD before and after matching

거주 여부 변수 (living inside capital area (Y/N))는 매우 불균형했다. 그리고 약간 불균형한 공변량들이 몇몇 존재했다.

처리군과 대조군의 표본 크기가 유사하므로 1:1 매칭을 진행하였다. 다음의 Table E.5는 매칭 후 두 그룹 간 공변량 균형을 나타낸다. 성향점수 매칭 (PSM), 공변량 균형 성향점수 매칭 (CBPSM), 그리고 성향점수 범위짜짓기 매칭 (caliper) 결과 총 420쌍이 매칭되었다. 매칭 전 두 그룹 간 단일 공변량이 매우 불균형했던 수도권 거주 여부 변수의 경우 여전히 표준화된 평균 차의 절댓값이 0.1 이상으로 나타났다. 성긴 정확 매칭 (CEM) 이용 시 본 분석과 동일하게 변수의 구간화를 시도하였을 때, 매칭이 불균형적이었다. 그러므로 가구 월 평균 소득 변수, 유아의 문해 및 언어능력 변수, 그리고 아버지 양육스트레스 변수를 4분위로 구간화하였다. 그 결과 성긴 정확 매칭 (CEM)은 모든 단일 공변량 평균 차이를 개선하였다. 하지만 상대적으로 적은 쌍 (152 개)을 매칭하였다. 카디널리티 (cardinality) 매칭은 모든 단일 공변량 평균을 균형적으로 만들었으며, 359쌍을 매칭하였다. 공변량 결합분포를 기준으로 두 그룹 간 공변량 균형을 평가하였을 때 성긴 정확 매칭 (CEM)의  $L_1$  통계량은 0.82로써 공변량 결합분포를 상대적으로 개선시켰다. 나머지 매칭 방법들은 공변량 결합분포를 개선하지 못했다.

다음의 Table E.6은 매칭 후 표본을 이용한 인과효과 추정 결과이다. 각 매칭 방법의 ATT 추정 결과는 일관적이지 않았지만, 모든 95% 신뢰구간의 하한과 상한 사이에 0을 포함하고 있으므로 두 변수 간 유의미한 인과효과는 존재하지 않는다.

Table D.1: Covariate balance before matching

Group		처리군	대조군	Two sample <i>t</i> -test <i>p</i> -value
$L_1$		0.95		
n		233	255	
Covariates Mean (Standard Deviation)	가구 월 평균 소득	447.04 (191.66)	440.47 (170.11)	0.42
	유아 문해 및 언어능력	54.90 (11.14)	52.67 (11.98)	0.33
	아버지 양육 스트레스	26.03 (6.69)	25.92 (6.45)	0.94
	수도권/비수도권_0	0.92 (0.50)	0.44 (0.46)	0.00
	수도권/비수도권_1	0.08 (0.46)	0.56 (0.50)	0.00
	아버지 최종학력_4	0.27 (0.43)	0.29 (0.45)	0.14
	아버지 최종학력_5	0.27 (0.43)	0.18 (0.40)	0.17
	아버지 최종학력_6	0.38 (0.49)	0.41 (0.50)	0.09
	아버지 최종학력_7	0.08 (0.27)	0.12 (0.30)	0.20
	어머니 최종학력_4	0.25 (0.46)	0.35 (0.45)	0.55
	어머니 최종학력_5	0.35 (0.47)	0.25 (0.44)	0.03
	어머니 최종학력_6	0.35 (0.48)	0.35 (0.49)	0.11
	어머니 최종학력_7	0.05 (0.22)	0.06 (0.22)	0.92

Table D.2: Covariate balance after matching

Method		PSM	CBPSM	Caliper	CEM	Cardinality
Matched pairs		233	233	233	40	129
$L_1$		0.95	0.95	0.94	0.88	0.96
Covariates Differences ( <i>p</i> -value)	가구 월 평균 소득	-0.01 (0.90)	-0.02 (0.79)	0.02 (0.87)	-0.05 (0.86)	0.00 (0.98)
	유아 문해 및 언어능력	0.10 (0.27)	0.13 (0.18)	0.18 (0.07)	-0.01 (0.96)	-0.01 (0.94)
	아버지 양육 스트레스	0.00 (0.98)	0.03 (0.78)	0.04 (0.69)	0.02 (0.91)	0.00 (0.97)
	수도권/비수도권_0	1.67 (0.00)	1.67 (0.00)	1.69 (0.00)	0.00 (1.00)	0.00 (1.00)
	수도권/비수도권_1	-1.67 (0.00)	-1.67 (0.00)	-1.69 (0.00)	0.00 (1.00)	0.00 (1.00)
	아버지 최종학력_4	0.02 (0.84)	0.01 (0.92)	-0.11 (0.27)	0.00 (1.00)	0.00 (1.00)
	아버지 최종학력_5	0.17 (0.05)	0.17 (0.05)	0.17 (0.05)	0.00 (1.00)	0.00 (1.00)
	아버지 최종학력_6	-0.08 (0.40)	-0.07 (0.45)	-0.06 (0.51)	0.00 (1.00)	0.00 (1.00)
	아버지 최종학력_7	-0.18 (0.09)	-0.18 (0.09)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력_4	-0.08 (0.40)	-0.08 (0.40)	-0.21 (0.03)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력_5	0.15 (0.09)	0.15 (0.09)	0.17 (0.06)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력_6	-0.05 (0.56)	-0.05 (0.56)	0.02 (0.85)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력_7	-0.06 (0.55)	-0.06 (0.55)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)

Table E.1: Results from regression adjustment

	ATE	SE	95% Lower Bound	95% Upper Bound
Regression Adjustment	-0.015	0.158	-0.325	0.295

Table E.2: Results from the weighting methods based on propensity score

	ATE	SE	95% Lower Bound	95% Upper Bound
IPW	-0.152	0.271	-0.672	0.389
SIPW	-0.038	0.178	-0.396	0.303
DR	-0.039	0.160	-0.358	0.268

Table E.3: Results from the weighting methods based on CBPS

	ATE	SE	95% Lower Bound	95% Upper Bound
IPW	-1.225	0.657	-2.988	-0.413
SIPW	-0.019	0.174	-0.364	0.318
DR	-0.032	0.159	-0.348	0.275

Table E.4: Covariate balance before matching

Group		처리군	대조군	Two sample <i>t</i> -test <i>p</i> -value
<i>L</i> <sub>1</sub>		0.95		
n		420	533	
Covariates Mean (Standard Deviation)	가구 월 평균 소득	447.76 (193.33)	437.45 (169.50)	0.39
	유아 문해 및 언어능력	54.04 (11.22)	53.22 (11.88)	0.27
	아버지 양육 스트레스	25.96 (6.72)	26.08 (6.44)	0.79
	수도권/비수도권_0	0.73 (0.44)	0.47 (0.50)	0.00
	수도권/비수도권_1	0.27 (0.44)	0.53 (0.50)	0.00
	아버지 최종학력_4	0.30 (0.46)	0.25 (0.43)	0.09
	아버지 최종학력_5	0.24 (0.43)	0.20 (0.40)	0.19
	아버지 최종학력_6	0.39 (0.49)	0.44 (0.50)	0.11
	아버지 최종학력_7	0.07 (0.26)	0.11 (0.31)	0.07
	어머니 최종학력_4	0.28 (0.45)	0.30 (0.46)	0.47
	어머니 최종학력_5	0.32 (0.47)	0.26 (0.44)	0.03
	어머니 최종학력_6	0.35 (0.48)	0.39 (0.49)	0.14
	어머니 최종학력_7	0.05 (0.22)	0.05 (0.22)	0.91

Table E.5: Covariate balance after matching

Method		PSM	CBPSM	Caliper	CEM	Cardinality
Matched pairs		420	420	420	152	359
$L_1$		0.95	0.95	0.94	0.82	0.94
Covariates Differences ( <i>p</i> -value)	가구 월 평균 소득	-0.02 (0.74)	-0.01 (0.94)	0.01 (0.93)	0.07 (0.56)	0.01 (0.90)
	유아 문해 및 언어능력	0.03 (0.68)	0.04 (0.62)	0.04 (0.57)	0.00 (0.99)	0.01 (0.90)
	아버지 양육 스트레스	-0.03 (0.66)	0.01 (0.93)	-0.03 (0.62)	0.04 (0.70)	0.01 (0.94)
	수도권/비수도권.0	0.31 (0.00)	0.31 (0.00)	0.31 (0.00)	0.00 (1.00)	0.01 (0.94)
	수도권/비수도권.1	-0.31 (0.00)	-0.31 (0.00)	-0.31 (0.00)	0.00 (1.00)	-0.01 (0.94)
	아버지 최종학력.4	0.01 (0.94)	0.05 (0.44)	0.01 (0.94)	0.00 (1.00)	0.01 (0.93)
	아버지 최종학력.5	0.03 (0.62)	0.01 (0.87)	0.02 (0.74)	0.00 (1.00)	0.00 (1.00)
	아버지 최종학력.6	-0.01 (0.89)	-0.06 (0.40)	-0.03 (0.67)	0.00 (1.00)	-0.01 (0.94)
	아버지 최종학력.7	-0.05 (0.52)	0.00 (1.00)	0.01 (0.89)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력.4	-0.05 (0.49)	-0.02 (0.82)	-0.10 (0.15)	0.00 (1.00)	-0.01 (0.93)
	어머니 최종학력.5	0.06 (0.37)	0.05 (0.46)	0.12 (0.08)	0.00 (1.00)	0.00 (1.00)
	어머니 최종학력.6	0.00 (1.00)	-0.03 (0.72)	-0.02 (0.77)	0.00 (1.00)	0.01 (0.94)
	어머니 최종학력.7	-0.03 (0.65)	-0.02 (0.76)	0.00 (1.00)	0.00 (1.00)	0.00 (1.00)

Table E.6: Results from the matching methods based on propensity score

	ATT	SE	95% Lower Bound	95% Upper Bound
PSM	0.151	0.179	-0.200	0.503
CBPSM	0.140	0.181	-0.210	0.490
Caliper	0.097	0.181	-0.254	0.448
CEM	-0.388	0.271	-0.957	0.181
Cardinality	-0.014	0.180	-0.382	0.354