



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0060603
(43) 공개일자 2019년06월03일

(51) 국제특허분류(Int. Cl.)

G06F 17/18 (2006.01)

(52) CPC특허분류

G06F 17/18 (2013.01)

(21) 출원번호 10-2017-0158959

(22) 출원일자 2017년11월24일

심사청구일자 2017년11월24일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

이명재

서울특별시 마포구 마포대로 195, 203동 2304호

김종원

서울특별시 용산구 장문로 27, 5동 102호

(74) 대리인

심경식, 홍성욱

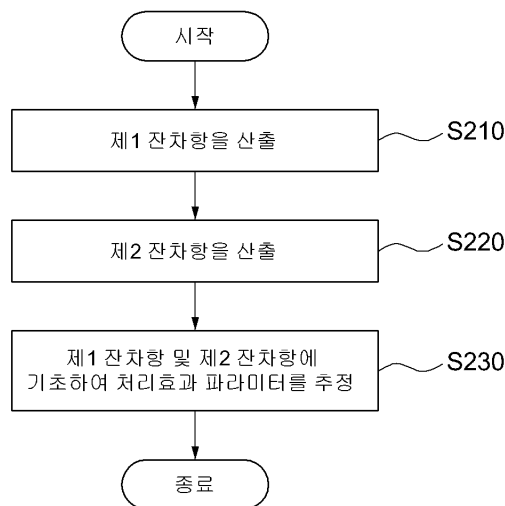
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법 및 장치**

(57) 요약

제1 잔차항 산출부가, 처리변수에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산출하는 단계, 제2 잔차항 산출부가, 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계 및 처리효과 추정부가, 제1 잔차항 및 제2 잔차항에 기초하여 처리변수가 결과변수에 미치는 처리효과 파라미터를 추정하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법 및 이를 위한 장치에 관한 것이다.

대표도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 1345265153

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 인문사회기초연구

연구과제명 처리효과를 위하여 성향점수 잔차항을 이용하는 최소자승추정법

기 여 율 1/1

주관기관 고려대학교

연구기간 2017.07.01 ~ 2018.06.30

명세서

청구범위

청구항 1

제1 잔차항 산출부가, 처리변수에 대한 처리변수모형 및 상기 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 상기 처리변수에서 상기 성향점수를 차감한 제1 잔차항을 산출하는 단계;

제2 잔차항 산출부가, 상기 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 상기 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 상기 결과변수에서 상기 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계; 및

처리효과 추정부가, 상기 제1 잔차항 및 상기 제2 잔차항에 기초하여 상기 처리변수가 상기 결과변수에 미치는 처리효과 파라미터를 추정하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법.

청구항 2

제1항에 있어서,

상기 제1 잔차항을 산출하는 단계는,

상기 처리변수모형 및 상기 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하는 단계;

상기 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하는 단계;

상기 샘플 우도 함수가 최대값이 되도록 상기 제1 파라미터 값을 결정하는 단계;

상기 제1 파라미터 값 및 상기 분포모형에 기초하여 상기 성향점수를 산출하는 단계; 및

상기 처리변수에서 상기 성향점수를 차감한 상기 제1 잔차항을 산출하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법.

청구항 3

제2항에 있어서,

상기 제1 파라미터 값을 결정하는 단계는,

상기 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 상기 로그화 샘플 우도 함수를 상기 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 상기 제1 파라미터 값으로 결정하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법.

청구항 4

제1항에 있어서,

상기 제2 잔차항을 산출하는 단계는,

상기 근사모형에 상기 제1 파라미터 값을 적용하여 제1 파라미터 적용 근사모형을 생성하는 단계;

상기 결과변수에서 상기 제1 파라미터 적용 근사모형을 차감하여 제2 오차항을 결정하는 단계;

상기 제2 오차항의 제곱이 최소가 되도록 하는 제2 파라미터의 값을 상기 제2 파라미터 값으로 결정하는 단계;

상기 제1 파라미터 적용 근사모형에 상기 제2 파라미터 값을 적용하여 상기 파라미터 적용 근사모형을 생성하는 단계; 및

상기 결과변수에서 상기 파라미터 적용 근사모형을 차감한 상기 제2 잔차항을 산출하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법.

청구항 5

제1항에 있어서,

상기 처리효과 파라미터를 추정하는 단계는,

상기 제1 잔차항과 상기 처리효과 파라미터를 곱한 뒤, 상기 제2 잔차항에서 상기 제1 잔차항과 상기 처리효과 파라미터를 곱한 결과를 차감하고, 상기 차감한 결과를 제공한 결과를 최소로 하는 상기 처리효과 파라미터 값을 추정하는 단계를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법.

청구항 6

처리변수에 대한 처리변수모형 및 상기 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 상기 처리변수에서 상기 성향점수를 차감한 제1 잔차항을 산출하는 제1 잔차항 산출부;

상기 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 상기 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 상기 결과변수에서 상기 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 제2 잔차항 산출부; 및

상기 제1 잔차항 및 상기 제2 잔차항에 기초하여 상기 처리변수가 상기 결과변수에 미치는 처리효과 파라미터를 추정하는 처리효과 추정부를 포함하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치.

청구항 7

제6항에 있어서,

상기 제1 잔차항 산출부는,

상기 처리변수모형 및 상기 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하고,

상기 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하고,

상기 샘플 우도 함수가 최대값이 되도록 상기 제1 파라미터 값을 결정하고,

상기 제1 파라미터 값 및 상기 분포모형에 기초하여 상기 성향점수를 산출하고,

상기 처리변수에서 상기 성향점수를 차감한 상기 제1 잔차항을 산출하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치.

청구항 8

제7항에 있어서,

상기 제1 잔차항 산출부는,

상기 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 상기 로그화 샘플 우도 함수를 상기 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 상기 제1 파라미터 값으로 결정하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치.

청구항 9

제6항에 있어서,

상기 제2 잔차항 산출부는,

상기 근사모형에 상기 제1 파라미터 값을 적용하여 제1 파라미터 적용 근사모형을 생성하고,

상기 결과변수에서 상기 제1 파라미터 적용 근사모형을 차감하여 제2 오차항을 결정하고,

상기 제2 오차항의 제공이 최소가 되도록 하는 제2 파라미터의 값을 상기 제2 파라미터 값으로 결정하고,

상기 제1 파라미터 적용 근사모형에 상기 제2 파라미터 값을 적용하여 상기 파라미터 적용 근사모형을 생성하고,

상기 결과변수에서 상기 파라미터 적용 근사모형을 차감한 상기 제2 잔차항을 산출하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치.

청구항 10

제6항에 있어서,

상기 처리효과 추정부는,

상기 제1 잔차항과 상기 처리효과 파라미터를 곱한 뒤, 상기 제2 잔차항에서 상기 제1 잔차항과 상기 처리효과 파라미터를 곱한 결과를 차감하고, 상기 차감한 결과를 제공한 결과를 최소로 하는 상기 처리효과 파라미터 값을 추정하는, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치.

발명의 설명

기술 분야

[0001] 본 발명은 성향점수 잔차항을 최소자승회귀분석에 활용하여 처리효과를 추정하기 위한 방법 및 장치에 관한 것이다.

배경 기술

[0002] 처리(Treatment)는 정부정책, 직업교육 프로그램 등과 같이 선택적으로 변경이 가능하며 결과에 영향을 줄 수 있는 변수들을 지칭하는 것으로, 예를 들어 처리가 “직업교육 프로그램”인 경우 결과는 “실업기간”을 의미할 수 있다.

[0003] 종래에 처리 효과(Treatment Effect)를 추정하기 위해서는 매칭(Matching) 방법을 활용하였으며, 매칭은 “처리”를 받는 사람들의 그룹과 “처리”를 받지 않는 사람들의 그룹에서 성격변수(H)가 비슷한 사람들을 뽑아 그들 사이에서 결과를 비교하여 처리의 효과를 추정하는 방법이다.

[0004] 이때, 성격변수란 결과에 영향을 미칠 수 있는 요소들로, 예컨대 성별, 학력 등과 같은 개인적인 요소가 포함될 수 있다.

[0005] 그러나, 종래의 매칭 방법은 활용 시에 “가장 비슷한” 것이 얼마나 비슷해야 하는지, 몇 명의 사람들을 비교해야 하는지 등의 선택 문제가 많고, 통계적 유의성을 판별하기 어려운 문제점이 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국 등록특허공보 제10-0557476호(2006.02.24.)

발명의 내용

해결하려는 과제

[0007] 본 발명의 목적은, 상기 문제점을 해결하기 위한 것으로, 처리변수에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산출하고, 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하고, 제1 잔차항 및 제2 잔차항에 기초하여 처리변수가 결과변수에 미치는 처리효과 파라미터를 추정하기 위함이다.

[0008] 본 발명이 해결하고자 하는 과제는 이상에서 언급한 과제(들)로 제한되지 않으며, 언급되지 않은 또 다른 과제(들)는 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0009] 상기한 목적을 달성하기 위하여 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법은 제1 잔차항 산출부가, 처리변수에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산출하는 단계, 제2 잔차항 산출부가, 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계 및, 제1 잔차항 및 제2 잔차항에 기초하여 처리변수가 결과변수에 미치는 처리효과 파라미터를 추정하는 단계를 포함한다.
- [0010] 예컨대, 제1 잔차항을 산출하는 단계는, 처리변수모형 및 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하는 단계, 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하는 단계, 샘플 우도 함수가 최대값이 되도록 제1 파라미터 값을 결정하는 단계, 제1 파라미터 값 및 분포모형에 기초하여 성향점수를 산출하는 단계 및 처리변수에서 성향점수를 차감한 제1 잔차항을 산출하는 단계를 포함한다.
- [0011] 예컨대, 제1 파라미터 값을 결정하는 단계는, 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 로그화 샘플 우도 함수를 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 제1 파라미터 값으로 결정하는 단계를 포함한다.
- [0012] 일 실시예에 따라, 제2 잔차항을 산출하는 단계는, 근사모형에 제1 파라미터 값을 적용하여 제1 파라미터 적용 근사모형을 생성하는 단계, 결과변수에서 제1 파라미터 적용 근사모형을 차감하여 제2 오차항을 결정하는 단계, 제2 오차항의 제곱이 최소가 되도록 하는 제2 파라미터의 값을 제2 파라미터 값으로 결정하는 단계, 제1 파라미터 적용 근사모형에 제2 파라미터 값을 적용하여 파라미터 적용 근사모형을 생성하는 단계 및 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계를 포함한다.
- [0013] 예를 들어, 처리효과 파라미터를 추정하는 단계는, 제1 잔차항과 처리효과 파라미터를 곱한 뒤, 제2 잔차항에서 제1 잔차항과 처리효과 파라미터를 곱한 결과를 차감하고, 차감한 결과를 제공한 결과를 최소로 하는 처리효과 파라미터 값을 추정하는 단계를 포함한다.
- [0014] 상기한 목적을 달성하기 위하여, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치는, 처리변수에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산출하는 제1 잔차항 산출부, 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 제2 잔차항 산출부 및 제1 잔차항 및 제2 잔차항에 기초하여 처리변수가 결과변수에 미치는 처리효과 파라미터를 추정하는 처리효과 추정부를 포함한다.
- [0015] 예를 들어, 제1 잔차항 산출부는, 처리변수모형 및 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하고, 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하고, 샘플 우도 함수가 최대값이 되도록 제1 파라미터 값을 결정하고, 제1 파라미터 값 및 분포모형에 기초하여 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산출한다.
- [0016] 예컨대, 제1 잔차항 산출부는, 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 로그화 샘플 우도 함수를 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 제1 파라미터 값으로 결정한다.
- [0017] 일 실시예에 따라, 제2 잔차항 산출부는, 근사모형에 제1 파라미터 값을 적용하여 제1 파라미터 적용 근사모형을 생성하고, 결과변수에서 제1 파라미터 적용 근사모형을 차감하여 제2 오차항을 결정하고, 제2 오차항의 제곱이 최소가 되도록 하는 제2 파라미터의 값을 제2 파라미터 값으로 결정하고, 결과변수에서 제2 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출한다.
- [0018] 일 실시예에 따라, 처리효과 추정부는, 제1 잔차항과 처리효과 파라미터를 곱한 뒤, 제2 잔차항에서 제1 잔차항과 처리효과 파라미터를 곱한 결과를 차감하고, 차감한 결과를 제공한 결과를 최소로 하는 상기 처리효과 파라미터 값을 추정한다.

발명의 효과

- [0019] 본 발명의 일 실시예에 따르면, 처리변수에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항에 대한 분포모형에 기초하여 제1 파라미터 값 및 성향점수를 산출하고, 처리변수에서 성향점수를 차감한 제1 잔차항을 산

출하고, 제1 파라미터 값에 기초하여 결과변수의 조건부평균값에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하고, 제1 잔차항 및 제2 잔차항에 기초하여 처리변수가 결과변수에 미치는 처리효과 파라미터를 추정함으로써, 사용하기가 간단하고 통계적 유의성을 쉽게 판별할 수 있는 처리효과 추정 방법 및 장치를 제공할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치를 설명하기 위한 구성도이다.

도 2는 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법을 설명하기 위한 순서도이다.

도 3은 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법에서 제1 잔차항을 산출하는 단계를 설명하기 위한 순서도이다.

도 4는 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법에서 제2 잔차항을 산출하는 단계를 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명의 기술적 사상을 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여, 본 발명의 가장 바람직한 실시예를 첨부 도면을 참조하여 설명하기로 한다. 우선 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면상에 표시되더라도 가능한 한 동일한 부호를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0022] 이하, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법 및 장치를 첨부된 도면을 참조하여 상세하게 설명하면 아래와 같다.

[0023] 우선, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법 및 장치에서 활용하는 처리(Treatment), 처리변수(D), 성격변수(H) 및 결과변수(Y)에 대하여 정의한다.

[0024] 예를 들어, 처리(Treatment)는 정부정책 및 직업교육 프로그램 등과 같이 선택적으로 변경이 가능하며 결과변수(Y)에 영향을 줄 수 있는 변수들을 의미할 수 있다.

[0025] 예컨대, 결과변수(Y)가 실업 기간인 경우 처리(Treatment)는 정부의 실업자 직업교육 정책을 의미할 수 있다.

[0026] 예를 들어, 처리변수(D)는 이항변수으로써, 특정 개인이 처리(Treatment)를 받았는지의 유무를 나타내며, D_i 가 1이라는 것의 의미는 분석 대상 그룹의 i 번째 개인은 처리(Treatment)를 받았음을 의미하며, D_j 가 0이라는 것의 의미는 분석 대상 그룹의 j 번째 개인은 처리(Treatment)를 받지 않았음을 의미할 수 있다.

[0027] 예를 들어, 처리(Treatment)가 정부의 실업자 직업교육 정책인 경우 D_i 가 1이라면 분석 대상 그룹의 i 번째 개인은 실업자 직업교육을 받았음을 의미하고, D_i 가 0이라면 분석 대상 그룹의 i 번째 개인은 실업자 직업교육을 받지 않았음을 의미할 수 있다.

[0028] 예컨대, 성격변수(H)는 결과변수(Y)에 영향을 줄 수 있는 변수들 중 통제 변수를 의미하며, 개인의 성격, 성별, 나이 등 결과변수(Y)에 영향을 줄 수 있는 통제변수를 의미할 수 있다.

[0029] 예를 들어, 결과변수(Y)는 처리변수(D)에 대한 종속변수으로써, 처리(Treatment)가 정부의 실업자 직업교육 정책인 경우, 결과변수(Y)는 “실업기간”을 의미할 수 있다.

[0030] 예컨대, 처리(Treatment)가 정부의 실업자 직업교육 정책이고, 처리변수(D)가 정부의 실업자 직업교육을 받았는지 여부이며, 결과변수(Y)가 실업기간을 의미한다면, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법 및 장치는 직업교육이 실업기간에 미치는 영향인 처리효과 파라미터(γ)를 추정할 수 있다.

- [0031] 이하, 도 1을 참조하여 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치(100)를 설명한다.
- [0032] 도 1은 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치를 설명하기 위한 구성도이다.
- [0033] 도 1에 도시된 바와 같이, 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치(100)는 제1 잔차항 산출부(110), 제2 잔차항 산출부(120) 및 처리효과 추정부(130)를 포함한다.
- [0034] 제1 잔차항 산출부(110)는 처리변수(D)에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항(v)에 대한 분포모형에 기초하여 제1 파라미터 값($\hat{\alpha}$) 및 성향점수($P(H_i)$)를 산출하고, 처리변수(D)에서 성향점수($P(H_i)$)를 차감한 제1 잔차항을 산출한다.
- [0035] 예컨대, 제1 잔차항 산출부(110)는, 처리변수모형 및 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하고, 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하고, 샘플 우도 함수가 최대값이 되도록 제1 파라미터 값($\hat{\alpha}$)을 결정하고, 제1 파라미터 값($\hat{\alpha}$) 및 분포모형에 기초하여 성향점수($P(H_i)$)를 산출하고, 처리변수(D)에서 상기 성향점수($P(H_i)$)를 차감한 제1 잔차항을 산출할 수 있다.
- [0036] 예컨대, 제1 잔차항 산출부(110)는, 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 로그화 샘플 우도 함수를 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 제1 파라미터 값($\hat{\alpha}$)으로 결정한다.
- [0037] 제2 잔차항 산출부(120)는 제1 파라미터 값($\hat{\alpha}$)에 기초하여 결과변수(Y)의 조건부평균값($E(Y_i | H_i' \hat{\alpha})$)에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값($\hat{\beta}$)을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수(Y)에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출한다.
- [0038] 예컨대, 제2 잔차항 산출부는, 근사모형에 제1 파라미터 값($\hat{\alpha}$)을 적용하여 제1 파라미터 적용 근사모형을 생성하고, 결과변수(Y)에서 제1 파라미터 적용 근사모형을 차감하여 제2 오차항을 결정하고, 제2 오차항의 제곱이 최소가 되도록 하는 제2 파라미터(β)의 값을 상기 제2 파라미터 값($\hat{\beta}$)으로 결정하고, 제1 파라미터 적용 근사모형에 제2 파라미터 값($\hat{\beta}$)을 적용하여 파라미터 적용 근사모형을 생성하고, 결과변수(Y)에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출한다.
- [0039] 처리효과 추정부(130)는 제1 잔차항 및 제2 잔차항에 기초하여 처리변수(D)가 결과변수(Y)에 미치는 처리효과 파라미터(γ)를 추정한다.
- [0040] 예컨대, 처리효과 추정부(130)는, 제1 잔차항과 처리효과 파라미터(γ)를 곱한 뒤, 제2 잔차항에서 제1 잔차항과 처리효과 파라미터(γ)를 곱한 결과를 차감하고, 그 제곱한 결과를 최소로 하는 처리효과 파라미터 값을 추정한다.
- [0041] 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 장치(100)에 대한 보다 구체적인 설명은 이하 도 2 내지 도 4를 참조하여 후술하도록 하며, 중복되는 설명은 생략한다.
- [0042] 이하, 도 2를 참조하여, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법을 설명한다.
- [0043] 도 2는 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법을 설명하기 위한 순서도이다.
- [0044] 도 2에 도시된 바와 같이, 본 발명의 실시예에 따른 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법은 제1 잔차항을 산출하는 단계(S210), 제2 잔차항을 산출하는 단계(S220) 및 제1 잔차항과 제2 잔차항에 기초하여 처리효과 파라미터를 추정하는 단계(S230)를 포함한다.

[0045] S210 단계는, 제1 잔차항 산출부(110)가, 처리변수(D)에 대한 처리변수모형 및 처리변수모형에 포함된 제1 오차항(v)에 대한 분포모형에 기초하여 제1 파라미터 값($\hat{\alpha}$) 및 성향점수($P(H_i)$)를 산출하고, 처리변수(D)에서 성향점수($P(H_i)$)를 차감한 제1 잔차항을 산출하는 단계를 의미할 수 있다.

[0046] 예를 들어, 처리변수(D)에 대한 처리변수모형은 회귀모형을 포함하되, 0 내지 1 사이의 범위를 가지는 이항 변수에 대하여 회귀 분석을 위해 활용되는 모형으로 기존에 공지된 각종 모형 중 적어도 하나를 의미할 수 있다.

[0047] 예를 들어, 처리변수(D)가 이항 변수가 아니라 0, 1, 2, ..., J의 값을 가지는 순서적(ordinal) 또는 범주형(categorical) 다항처리(multiple treatment)인 경우, 회귀 모형은 순서적(ordinal) 또는 범주형(categorical) 다항처리(multiple treatment)를 위한 각종 모형 중 적어도 하나를 의미할 수도 있다.

[0048] 다시 말해, 처리변수모형은 처리변수(D)를 성격변수(H)에 대해 회귀 분석하기 위한 모형으로, 처리(Treatment)가 정부의 실업자 직업교육 정책인 경우, 처리변수(D)는 직업교육 여부를 의미하고, 성격변수(H)는 개인의 나이, 성별 등 각종 변수를 의미할 수 있다.

[0049] 예를 들어, 처리변수모형이 로지스틱 회귀 분석(Logistic Regression)을 위한 모형인 경우, 처리변수모형은 아래 수학적 식 1과 같이 생성될 수 있으나, 본 발명은 이에 한정되지 않으며, 본 발명은 프로빗(Probit)을 비롯하여 다른 처리변수모형을 활용할 수 있다.

[0050] [수학적 식 1]

$$D_i^* = H_i' \alpha + v_i,$$

$$D_i = 1[D_i^* \geq 0] = 1[H_i' \alpha + v_i \geq 0]$$

[0051]

[0052] 이때, D_i^* 는 i번째 개인에 대한 잠재변수(latent variable)로 연속 변수의 성질을 나타내며, H_i' 는 i번째 개인에 대한 적어도 하나 이상의 성격변수(H)를 포함하는 성격벡터, α 는 제1 파라미터 벡터, v_i 는 i번째 개인에 대한 제1 오차항, D_i 는 i번째 개인에 대한 처리변수, i는 1 내지 N의 범위를 가지는 주어진 자료 내에서 i번째 개인, $H_i' \alpha$ 은 잠재변수(D_i^*)의 회귀 모형을 의미한다.

[0053] 예를 들어, 성격벡터(H_i')가 [M by 1] 벡터인 경우, 제1 파라미터 벡터(α)는 [M by 1] 벡터를 의미할 수 있으며, 이 경우, 벡터의 내적 연산에 의하여 $H_i' \alpha$ 는 스칼라(scalar) 값을 가질 수 있다.

[0054] 여기서, 잠재변수(D_i^*)가 0 보다 크거나 같은 경우 처리변수(D_i)는 1이 되고, 잠재변수(D_i^*)가 0 보다 작은 경우 처리변수(D_i)는 0이 될 수 있다.

[0055] 한편, 제1 파라미터 벡터(α)는 로지스틱 회귀 분석(Logistic Regression)에서 추정하고자 하는 모수(parameter)로서 개인의 성격을 의미하는 성격변수(H_i)가 잠재변수(D_i)에 미치는 영향의 정도를 나타낸다.

[0056] 한편, 제1 오차항(error term, v_i)은 잠재변수(D_i^*)에서 성격변수(H_i)가 설명하지 못하는 부분을 나타낸다.

[0057] 예를 들어, 제1 오차항(v_i)이 로지스틱 분포를 따르는 경우, 제1 오차항(v_i)에 대한 분포모형은 아래 수학적 식 2와 같을 수 있으나, 본 발명은 제1 오차항(v_i)이 로지스틱 분포를 따르는 경우에 한정되지 않고, 제1 오차항(v_i)은 로지스틱 분포 외에도 정규분포를 비롯한 각종 분포를 따를 수도 있다.

[0058] [수학적 식 2]

$$G(v) = \frac{e^v}{1 + e^v}$$

[0059]

- [0060] 이때, $G(v_i)$ 는 제1 오차항(v_i)의 분포함수를 의미하며, 이러한 로지스틱 분포는 0을 중심으로 대칭이다.
- [0061] 이때, 수학적 1에 기재된 모형 및 수학적 2에 기재된 분포모형은 일 실시예로써, 본 발명은 수학적 1 및 수학적 2로 한정되지 않는다.
- [0062] 이제, 도 3을 참조하여 S210 단계를 계속 설명한다.
- [0063] 도 3은 본 발명의 실시예에 따른, 성향점수 잔차항을 최소자승회귀분석에 활용하는 처리효과 추정 방법에서 제1 잔차항을 산출하는 단계를 설명하기 위한 순서도이다.
- [0064] 도 3에 도시된 바와 같이, S210 단계는 개인별 우도 함수를 결정하는 단계(S211), 샘플 우도 함수를 산출하는 단계(S213), 제1 파라미터 값을 결정하는 단계(S215), 성향점수를 산출하는 단계(S217) 및 제1 잔차항을 산출하는 단계를 포함한다.
- [0065] S211 단계는, 처리변수모형 및 분포모형에 대응되는 개인별 우도 함수(likelihood function)를 결정하는 단계를 의미할 수 있다.
- [0066] 예를 들어, 처리변수모형이 상술한 수학적 1과 같고 분포모형이 상술한 수학적 2와 같은 경우, 개인별 우도 함수는 아래 수학적 3과 같이 산출될 수 있다.
- [0067] 한편, 처리변수모형 및 분포모형이 상술한 수학적 1 및 수학적 2와 다른 경우, 개인별 우도 함수는 기 공지된 최대 우도 추정 알고리즘(Maximum Likelihood Estimator)에서 결정하는 각종 우도 함수 결정 방식을 따를 수 있으며, 본 발명에 따른 개인별 우도 함수는 아래 수학적 3에 한정되지 않는다.
- [0068] [수학적 3]

$$\begin{aligned}
 & P(D_i = 1|H_i)^{D_i}P(D_i = 0|H_i)^{1-D_i} \\
 & = P(D_i^* \geq 0|H_i)^{D_i}P(D_i^* < 0|H_i)^{1-D_i} \\
 & = P(v_i \geq -H_i'\alpha|H_i)^{D_i}P(v_i < -H_i'\alpha|H_i)^{1-D_i} \\
 & = P(-v_i \leq H_i'\alpha|H_i)^{D_i}G(-H_i'\alpha)^{1-D_i} \\
 & = G(H_i'\alpha)^{D_i}\{1 - G(H_i'\alpha)\}^{1-D_i}
 \end{aligned}$$

- [0069]
- [0070] 이때, $G()$ 는 제1 오차항(v_i)의 분포함수, H_i' 는 i 번째 개인에 대한 적어도 하나 이상의 성격변수(H)를 포함하는 성격벡터, α 는 제1 파라미터 벡터, D_i 는 i 번째 개인에 대한 처리변수를 의미한다.
- [0071] S213 단계는, 개인별 우도 함수를 모두 곱하여 샘플 우도 함수(sample likelihood function)를 산출하는 단계를 의미할 수 있다.
- [0072] 이때, 샘플 우도 함수(sample likelihood function)는 아래 수학적 4에 의해 결정된다.
- [0073] [수학적 4]

$$L(\alpha) = \prod_{i=1}^N G(H_i'\alpha)^{D_i}\{1 - G(H_i'\alpha)\}^{1-D_i}$$

- [0074]
- [0075] 이때, $L(\alpha)$ 는 샘플 우도 함수를 의미한다.
- [0076] S215 단계는, 샘플 우도 함수가 최대값이 되도록 제1 파라미터(α) 값을 결정하는 단계를 의미할 수 있다.
- [0077] 예를 들어, S215 단계는, 샘플 우도 함수에 로그를 취하여 로그화 샘플 우도 함수를 산출하고, 로그화 샘플 우도 함수를 제1 파라미터에 대해 미분한 결과가 0이 되도록 하는 값을 제1 파라미터 값으로 결정하는 단계를 포함할 수 있다.

[0078] 예를 들어, 상술한 S211 단계 내지 S215 단계는 일종의 최대 우도 추정 알고리즘(Maximum Likelihood Estimator)을 의미할 수 있으며, 처리변수(D)모형의 오차항이 로지스틱 분포를 따르는 경우 로지스틱 분포에 기반한 이항변수에 대한 최대 우도 추정 알고리즘(Maximum Likelihood Estimator)을 의미할 수 있고, 처리변수(D)모형의 오차항이 정규 분포를 따르는 경우 정규 분포에 기반한 이항변수에 대한 최대 우도 추정 알고리즘(Maximum Likelihood Estimator)을 의미할 수 있다.

[0079] 이때, 샘플 우도 함수가 최대값이 되도록 하는 제1 파라미터(α)의 값인 제1 파라미터 값($\hat{\alpha}$)을 추정하는 과정은 로지스틱 회귀 분석(logistic regression)결과를 의미할 수 있으나, 본 발명은 이에 한정되지 않는다.

[0080] 예컨대, S215 단계에서 활용되는 로그화 샘플 우도 함수는 아래 수학적식 5와 같이 표현될 수 있다.

[0081] [수학적식 5]

$$\ln L(\alpha) = \sum_{i=1}^N [D_i \ln G(H'_i \alpha) + (1 - D_i) \ln \{1 - G(H'_i \alpha)\}]$$

[0082]

[0083] 한편, 로그화 샘플 우도 함수를 제1 파라미터(α)에 대해 미분한 결과는 아래 수학적식 6과 같이 표현될 수 있다.

[0084] [수학적식 6]

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \sum_{i=1}^N \frac{\{D_i - G(H'_i \alpha)\} H'_i G'(H'_i \alpha)}{G(H'_i \alpha) \{1 - G(H'_i \alpha)\}}$$

[0085]

[0086] 이때, 수학적식 6을 0으로 만드는 제1 파라미터(α)의 값은 샘플 우도 함수가 최대값이 되도록 하는 제1 파라미터 값($\hat{\alpha}$)을 의미할 수 있다.

[0087] S217 단계는, 제1 파라미터 값($\hat{\alpha}$) 및 처리변수모형에 기초하여 성향점수를 산출하는 단계를 의미할 수 있다.

[0088] 예를 들어, S217 단계는 상술한 수학적식 2의 분포함수에 성격벡터(H_i')와 샘플 우도 함수가 최대값이 되도록 하는 제1 파라미터 값($\hat{\alpha}$)을 대입하여, 성향점수를 산출하는 단계를 의미할 수 있으며 이러한 과정은 하기의 수학적식 7과 같을 수 있으나, 본 발명은 이에 한정되지 않는다.

[0089] [수학적식 7]

$$P(H_i) = P(D = 1 | H_i) = G(H'_i \hat{\alpha}) = \frac{e^{H'_i \hat{\alpha}}}{1 + e^{H'_i \hat{\alpha}}}$$

[0090]

[0091] 이때, $P(H_i)$ 는 성향 점수, H_i 는 i 번째 개인에 대한 성격변수, $G(\cdot)$ 는 제1 오차항(v_i)의 분포함수, H_i' 는 i 번째 개인에 대한 성격벡터, $\hat{\alpha}$ 는 샘플 우도 함수가 최대값이 되도록 하는 제1 파라미터 값을 의미한다.

[0092] 예를 들어, 성향 점수(Propensity Score)는 각 개인이 처리(Treatment)를 받았을 확률(처리변수(D)가 1인 확률)로 정의할 수 있다.

[0093] 예를 들어, 처리(Treatment)가 정부의 실업자 직업교육 정책인 경우, 성향 점수(Propensity Score)는 각 개인이 직업 교육을 받을 확률인 $P(D = 1 | H_i)$ 로 정의할 수 있다.

[0094] S219 단계는, 각각의 개인에 대한 처리변수(D_i)에서 각 개인에 대한 성향점수($P(H_i)$)를 차감한 제1 잔차항($D_i - P(H_i)$)을 산출하는 단계를 의미할 수 있다.

[0095] 예를 들어, S211 내지 S219 단계에서는 상술한 로지스틱 회귀 분석 외에도 프로빗(Probit) 회귀 분석 방식을 사용할 수도 있으며, D를 H에 회귀분석하는 최소자승법 및 기타 H를 예측하기 위하여 P(H)를 구하는 각종 방법이 활용가능하며, 상술한 로지스틱 회귀 분석 방식은 일 실시예에 불과하다.

[0096] 다시 도 2를 참조하여, S220 단계를 계속 설명한다.

[0097] S220 단계는, 제2 잔차항 산출부(120)가, 제1 파라미터 값($\hat{\alpha}$)에 기초하여 결과변수(Y)의 조건부평균값($E(Y_i | H_i' \hat{\alpha})$)에 대한 근사모형에 포함된 적어도 하나의 제2 파라미터 값($\hat{\beta}$)을 산출하여 근사모형에 적용함으로써 파라미터 적용 근사모형을 생성하고, 결과변수(Y)에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계를 의미할 수 있다.

[0098] 이때, 근사모형은 결과변수(Y)의 조건부평균값을 근사하기 위해 활용되는 각종 모형을 의미할 수 있으며, 본 발명은 특정 근사모형으로 한정되지 않는다.

[0099] 예를 들어, 근사모형이 다항식의 형태인 경우, 근사모형은 아래 수학식 8과 같이 도식될 수 있으나, 본 발명은 이에 한정되지 않는다.

[0100] [수학식 8]

$$E(Y_i | H_i' \alpha) = \beta_0 + \beta_1(H_i' \alpha) + \beta_2(H_i' \alpha)^2 + \dots + \beta_n(H_i' \alpha)^n$$

[0102] 이때, $E(Y_i | H_i' \alpha)$ 는 결과변수(Y)의 조건부평균값, $\beta_0, \beta_1, \beta_2, \beta_n$ 각각은 제2 파라미터, Y_i 는 i번째 개인에 대한 결과변수, H_i 는 i번째 개인에 대한 성격벡터, α 는 제1 파라미터, n은 근사모형을 생성하기 위해 결정된 임의의 자연수를 의미한다.

[0103] 예를 들어, S220 단계는 결과변수(Y)를 1, $H_i' \alpha$, $(H_i' \alpha)^2, \dots, (H_i' \alpha)^n$ 에 회귀분석하는 최소자승법(Least Squares Estimation)을 적용하는 단계를 의미할 수 있으나, 본 발명은 이에 한정되지 않는다.

[0104] 이하, 도 4를 참조하여, S220 단계에 대하여 보다 구체적으로 설명한다.

[0105] 도 4에 도식된 바와 같이, S220 단계는 제1 파라미터 적용 근사모형을 생성하는 단계(S221), 제2 오차항을 결정하는 단계(S223), 제2 파라미터 값을 결정하는 단계(S225), 파라미터 적용 근사모형을 생성하는 단계(S227) 및 제2 잔차항을 산출하는 단계(S229)를 포함한다.

[0106] S221 단계는, 근사모형에 제1 파라미터 값($\hat{\alpha}$)을 적용하여 제1 파라미터 적용 근사모형을 생성하는 단계를 의미할 수 있다.

[0107] 이때, 근사모형이 상술한 수학식 8과 같은 경우, 제1 파라미터 적용 근사모형은 아래 수학식 9와 같이 생성될 수 있으나, 본 발명은 이에 한정되지 않는다.

[0108] [수학식 9]

$$E(Y_i | H_i' \hat{\alpha}) = \beta_0 + \beta_1(H_i' \hat{\alpha}) + \beta_2(H_i' \hat{\alpha})^2 + \dots + \beta_n(H_i' \hat{\alpha})^n$$

[0110] 이때, $E(Y_i | H_i' \hat{\alpha})$ 는 제1 파라미터 값($\hat{\alpha}$)을 적용했을 때, 결과변수(Y)의 조건부 평균값, $\beta_0, \beta_1, \beta_2, \beta_n$ 각각은 제2 파라미터, Y_i 는 i번째 개인에 대한 결과변수, H_i 는 i번째 개인에 대한 성격벡터, $\hat{\alpha}$ 는 샘플 우도 함수가 최대값이 되도록 하는 제1 파라미터 값, n은 근사모형을 생성하기 위해 결정된 임의의 자연수를 의미한다.

[0111] S223 단계는 결과변수(Y)에서 제1 파라미터 적용 근사모형을 차감하여 제2 오차항(u_i)을 결정하는 단계를 의미할 수 있다.

[0112] 예를 들어, S223 단계는 각각의 개인별 결과변수(Y_i)에서 제1 파라미터 적용 근사모형을 차감하여 제2 오차항

(u_i)을 결정하는 단계를 의미할 수 있다.

[0113] 예컨대, 제1 파라미터 적용 근사모형이 상술한 수학적 식 9와 같은 경우, 제2 오차항(u_i)은 아래 수학적 식 10과 같이 산출될 수 있으나, 본 발명은 이에 한정되지 않는다.

[0114] [수학적 식 10]

[0115]
$$u_i = Y_i - E(Y_i | H_i' \hat{\alpha})$$

[0116] 이때, u_i 는 각각의 개인별 오차항, Y_i 는 각각의 개인별 결과변수, $E(Y_i | H_i' \hat{\alpha})$ 는 각각의 개인별 제1 파라미터 적용 근사모형을 의미한다.

[0117] S225 단계는, 제2 오차항(u_i)의 제곱이 최소가 되도록 하는 제2 파라미터(β)의 값을 제2 파라미터 값($\hat{\beta}$)으로 결정하는 단계를 의미할 수 있다.

[0118] 보다 구체적으로, S225 단계는 각각의 개인별 제2 오차항(u_i)의 제곱의 총합이 최소가 되도록 하는 제2 파라미터(β)의 값을 제2 파라미터 값($\hat{\beta}$)으로 결정하는 단계를 의미할 수 있다.

[0119] 상술한 수학적 식 9에서, $\beta_0, \beta_1, \beta_2, \beta_n$ 각각은 최소자승법에서 추정하고자 하는 모수(parameter)로써, 1, $H' \alpha, (H' \alpha)^2, (H' \alpha)^n$ 이 결과변수(Y_i)에 미치는 영향의 정도를 나타내고, 제2 오차항(u_i)은 결과변수(Y_i)가 1, $H' \alpha, (H' \alpha)^2, (H' \alpha)^n$ 에 의하여 설명되지 않는 나머지 부분을 의미할 수 있다.

[0120] 예를 들어, S225 단계에서 활용되는 개인별 제2 오차항(u_i)의 제곱의 총합은 아래 수학적 식 11과 같이 나타낼 수 있으나, 본 발명은 이에 한정되지 않는다.

[0121] [수학적 식 11]

[0122]
$$\sum_{i=1}^N u_i^2 = \sum_{i=1}^N (Y_i - \{\beta_0 + \beta_1 \cdot (H_i' \hat{\alpha}) + \beta_2 \cdot (H_i' \hat{\alpha})^2 + \dots + \beta_n \cdot (H_i' \hat{\alpha})^n\})^2$$

[0123] 이때, 상술한 수학적 식 11이 최소가 되도록 하는 제2 파라미터($\beta_0, \beta_1, \beta_2, \beta_n$) 각각을 구하는 방법은 기존에 공지된 최소자승법과 동일하므로 그 구체적인 설명을 생략하며, 상술한 수학적 식 11이 최소가 되도록 하는 제2 파라미터 값은 $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_n$ 과 같이 표현될 수 있다.

[0124] 예를 들어, 상술한 수학적 식 11이 최소가 되도록 하는 적어도 하나의 제2 파라미터 값을 산출하기 위하여, 상술한 수학적 식을 제2 파라미터($\beta_0, \beta_1, \beta_2, \beta_n$) 각각에 대하여 미분한 뒤, 도출된 1계조건을 모두 0으로 만드는 값을 제2 파라미터 값($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_n$)으로 산출할 수 있으나, 본 발명은 이에 한정되지 않는다.

[0125] S227 단계는, 제1 파라미터 적용 근사모형에 제2 파라미터 값을 적용하여 파라미터 적용 근사모형을 생성하는 단계를 의미할 수 있다.

[0126] 예를 들어, 제1 파라미터 적용 근사모형이 상기한 수학적 식 9와 같고 제2 파라미터 값이 ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_n$)인 경우, 파라미터 적용 근사모형은 아래 수학적 식 12와 같이 산출될 수 있으나, 본 발명은 이에 한정되지 않는다.

[0127] [수학적 식 12]

[0128]
$$\hat{\beta}_0 + \hat{\beta}_1(H_i' \hat{\alpha}) + \hat{\beta}_2(H_i' \hat{\alpha})^2 + \dots + \hat{\beta}_n(H_i' \hat{\alpha})^n$$

[0129] S229 단계는, 결과변수(Y_i)에서 파라미터 적용 근사모형을 차감한 제2 잔차항을 산출하는 단계를 의미할 수 있다.

[0130] 이때, 제2 잔차항은 아래 수학적 식 13과 같이 산출될 수 있다.

[0131] [수학적 식 13]

$$Y_i - \{\hat{\beta}_0 + \hat{\beta}_1(H_i' \hat{\alpha}) + \hat{\beta}_2(H_i' \hat{\alpha})^2 + \dots + \hat{\beta}_n(H_i' \hat{\alpha})^n\}$$

[0132]

[0133] 계속, 도 2를 참조하여 S230 단계를 설명한다.

[0134] S230 단계는, 처리효과 추정부(130)가, 제1 잔차항 및 제2 잔차항에 기초하여 처리변수(D)가 결과변수(Y)에 미치는 처리효과 파라미터(γ)를 추정하는 단계를 의미할 수 있다.

[0135] 예를 들어, S230 단계는 제1 잔차항과 처리효과 파라미터를 곱한 뒤, 제2 잔차항에서 제1 잔차항과 처리효과 파라미터를 곱한 결과를 차감하고, 이를 제곱한 결과를 최소로 하는 상기 처리효과 파라미터 값을 추정하는 단계를 포함할 수 있다.

[0136] 예컨대, S230 단계는 제2 잔차항을 제1 잔차항에 단순 회귀 분석하는 최소자승법을 적용함으로써, 처리효과 파라미터(γ)를 추정할 수 있으며, 그 처리효과 파라미터(γ)는 성격변수(H)를 통제한 상태에서 처리변수(D)가 결과변수(Y)에 미치는 효과를 의미할 수 있다.

[0137] 예컨대, S230 단계는 결과변수(Y)의 잔차항인 제2 잔차항을 종속 변수로하고 처리변수(D)의 잔차항인 제2 잔차항을 설명 변수로 하는 최소자승법을 의미할 수 도 있다.

[0138] 이때, S230 단계는 아래 수학적 식 14와 같이 표현될 수 있으나, 본 발명은 이에 한정되지 않는다.

[0139] [수학적 식 14]

$$\sum_{i=1}^N [B_i - \gamma A_i]^2$$

[0140]

[0141] 이때 B_i 는 i 번째 개인에 대한 제2 잔차항, A_i 는 i 번째 개인에 대한 제1 잔차항, γ 는 처리효과 파라미터를 의미한다.

[0142] 예를 들어, S230 단계는 상술한 수학적 식 14를 최소로 하는 처리효과 파라미터(γ)의 값을 추정하는 단계를 의미할 수 있다.

[0143] 이때, 처리효과 파라미터(γ)의 의미는 아래 수학적 식 15와 같이 설명될 수 있다.

[0144] [수학적 식 15]

$$\gamma = \frac{1}{N} \sum_{i=1}^N [Y_i^1 - Y_i^0]$$

[0145]

[0146] 이때, γ 는 처리효과 파라미터, N 은 분석 대상 그룹에 포함되는 전체 개인의 수, Y_i^1 은 i 번째 개인에 대하여 처리(treatment)가 수행되었을 때의 잠재적 결과변수, Y_i^0 은 i 번째 개인에 대하여 처리(treatment)가 수행되지 않았을 때의 잠재적 결과변수를 의미한다.

[0147] 예를 들어, 처리(Treatment)가 정부의 실업자 직업교육 정책인 경우, Y_i^1 은 i 번째 개인이 직업교육을 받았을 때의 실업기간, Y_i^0 은 i 번째 개인이 직업교육을 받지 않았을 때의 실업기간을 의미한다.

[0148] 상술한 S230 단계의 예시는 최소자승법을 활용하는 것으로 작성되었으나, 일 실시예에 따르면 $P(H) \times \{1 -$

$P(H)\}^{-0.5}$ 를 활용하는 가중최소자승법(weighted least squares estimator)를 활용할 수도 있으며, 이 경우, $P(H)$ 가 0 또는 1에 가까워 가중치가 무한대에 가까워지는 자료들을 제외하는 것도 가능하다.

[0149] 예를 들어, 본 발명의 실시예에 따른 회귀분석을 활용하는 처리효과 추정 방법 및 장치는, 처리변수(D)가 이항변수인 경우로 설명되었으나, 처리변수(D)가 이항변수가 아니라, 0,1,2,...,J의 값을 갖는 순서적(ordinal) 혹은 범주형(categorical) 다항처리(multiple treatment)인 경우, 이항터미변수 D_j 를 $D=j$ 의 경우로 정의(즉, $D=j$ 이면 $D_j=1$ 이며 아니면 $D_j=0$)하여 적용될 수도 있다.

[0150] 이 경우, S210 단계에서 D가 1,2,...,J가 될 확률 $P_1(H), P_2(H), \dots, P_J(H)$ 를 다양한 확률기반 연산 기법으로 구할 수 있으며, 최종적으로 수학적 식 13과 유사한 종속변수를 $D_1-P_1(X), \dots, D_J-P_J(X)$ 에 회귀분석 하는 최소자승법을 적용하면, 그 추정계수들은 $D=1,2,\dots,J$ 가 각각 Y에 주는 효과가 될 수 있으며, 본 발명은 처리변수(D)가 이항변수인 경우로 한정되지 않는다.

[0151] 이상에서 본 발명에 따른 바람직한 실시예에 대해 설명하였으나, 다양한 형태로 변형이 가능하며, 본 기술분야에서 통상의 지식을 가진 자라면 본 발명의 특허청구범위를 벗어남이 없이 다양한 변형예 및 수정예를 실시할 수 있을 것으로 이해된다.

부호의 설명

[0152] 100: 회귀분석을 활용하는 처리효과 추정 장치

110: 제1 잔차항 산출부

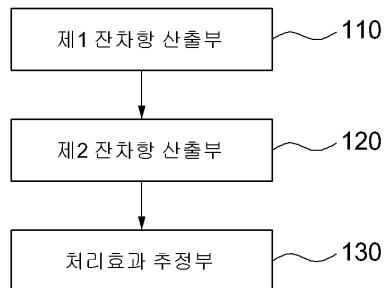
120: 제2 잔차항 산출부

130: 처리효과 추정부

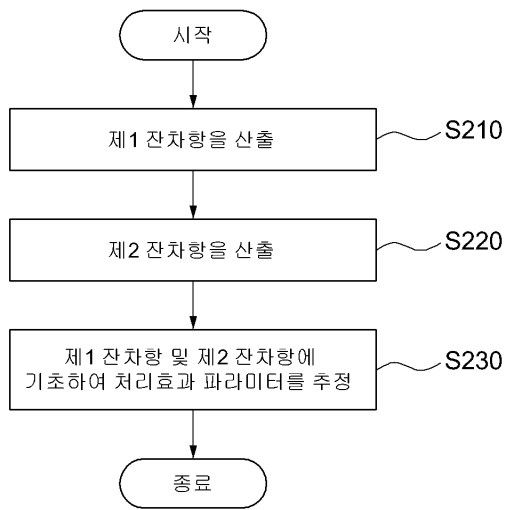
도면

도면1

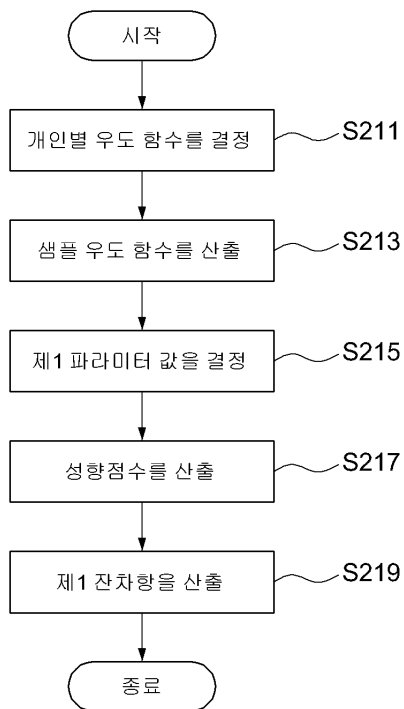
100



도면2



도면3



도면4

