노 동 정 책 연 구
2021. 제21권 제3호 pp.89~113
한 국 노 동 연 구 원
http://doi.org/10.22914/jlp.2021.21.3.004

연구논문

국민건강보험 빅데이터를 이용한 출산율과 환경에 대한 연구 : 예비적 분석*

김 준 일**

본 연구는 새로운 실증적 접근을 이용해 출산율과 환경의 관계에 대해 분석하는 것을 그 목표로 한다. 출산을 결정하는 요인으로서 환경에 대한 많은 연구가 존재하지만, 기존의 접근은 출산에 영향을 미치는 장소 특수적인(place-specific) 환경의 효과(Place Effect), 개인 간 이질성으로부터 나오는 선별효과(Selection Effect)를 구분하지 않고 실증적 분석을 하는 한계를 지닌다. 이에 국민건강보험 빅데이터를 이용하는 준실험적 연구방법론을 고안, 이주자(Mover)의 출생아 수가 환경에 어떤 영향을 받는지 살펴봄으로써 좀 더 효과적인 저출산정책을 수립하는 데 기여하고자 하였다. 실증 결과, 환경은 출산율 증가에 유의미한 영향을 미치기는 하지만 출산에 좋은 환경이 초기에 출산율을 상승시키는 부분 중 약 65% 정도는 선별효과에 의한, 즉 출산 가능성이 높은 세대를 모으는 효과에 의한 것이고 실제로 출산율을 상승시키는 효과는 그 이후에 서서히 나타나 6년 정도에 가장 정점에 도달했다가 이후 정체된다는 것을 알 수 있었다. 이 연구는 이후 구체적인 정책분석을 위한 예비적 분석으로서의 의미를 지닌다.

핵심용어 : 출산율, 선별효과, 장소효과, 노출효과, 준실험적 방법, 국민건강보험 빅데이터

논문접수일 : 2021년 4월 5일, 심사의뢰일 : 2021년 4월 12일, 심사완료일 : 2021년 5월 27일

* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구이다. (NRF-2018S1A5A8027705). 또한 이 연구는 국민건강보험공단의 국민건강정보자료(NHIS-2019-1-360)를 활용한 것으로, 연구의 결과는 국민건강보험공단과 관련이 없다.

** 목원대학교 글로벌경제학과 조교수(kjoonil@mokwon.ac.kr)

I. 서 론

지난 15여 년간 한국의 중앙정부와 지방자치단체는 막대한 자원을 동원하여 다양한 형태의 출산장려정책을 사용해왔다. 하지만 한국의 합계 출산율(Total Fertility Rate : TFR)은 지속적으로 하락하여 2020년도에는 0.84를 기록하였고 이는 세계 최하위 수준이다. 낮은 출산율과 이에 따른 고령화는 한국의 사회·경제·문화 전 분야에 걸쳐 심각한 변화를 초래할 것이기 때문에 출산율에 대한 사회과학적인 분석과 적절한 정책은 중요한 문제이다.

본 연구는 출산율과 관련된 두 가지 의문에서 출발하였다. 첫 번째 의문은 출산율의 지역적 차이는 왜 발생하는가이다. 주지하다시피 TFR은 지역별로 꽤 많은 차이가 존재한다. 예를 들어 세종시의 TFR은 2019년 1.47인 반면 서울은 0.72에 불과하다. 이 차이는 어디서 기인한 것일까? 출산율의 결정요인에 관한 기존의 연구들, 예를 들어, 천현숙 외(2012), 강정구·마강래(2016), 서문희 외(2016), 조영태 외(2018) 등의 실증분석은 출산율을 결정하는 요소로 크게 자산이나 고용형태 등 개인특성과 주택가격이나 출산장려금 등 지역적 환경특성으로 나누고 있다고 볼 수 있다. 하지만 개인들의 지역이동이 빈번한 상황에서 이 두 특성에 따른 효과를 구분하여 실증할 수 있을까? 특히 분석데이터가 여러 기간 동안의 지역별 데이터라고 한다면 두 효과는 혼재되어 나타날 가능성이 있다.

출산율과 관련된 두 번째 의문은 저출산정책이 얼마나 효과적인 정책이었는가이다. 특히 출산장려금 등 임신과 출산에 대한 각종 현금 및 비현금성 지역특수적 정책들이 특정지역의 출산율을 상승시키는 효과가 있었는가? 특히 앞서 언급하였던 것처럼 출산을 결정하는 요소를 두 가지 요인으로 구분하였을 때 혹시 정책의 효과들에는 단지 다른 지역의 출산율을 낮추며 (출산확률이 높은) 예비적 출산인구를 모으는(선별하는) 효과가 혼재되어 있지 않을까?[1] 출산율

1) 예를 들어 출산장려금정책의 효과와 관련된 최근의 논문들, 예를 들어 김우영·이정만(2018), 허만형(2020) 등은 이 정책이 지역의 출산율을 상승시킬 수도 있지만 출산가능인

의 변화가 단기적으로 가능하지 않다는 것을 보여주는 여러 선진국들의 사례를 고려할 때 더욱 그러하다. 그렇다면 지자체의 무분별한 정책들은 소위 인근궁핍화만을 초래할 뿐이지 않을까. 최근 신문기사[2]에도 정책의 이러한 제로섬게임적 성격이 지적된 바 있다.

이러한 질문들에 답하기 위해서는 특히 환경이 출산율에 미치는 영향에 대한 엄밀한 실증분석이 필요한데 그간 많은 연구에도 불구하고 기본적으로 출산이라는 것이 다양한 환경에 영향을 받는 매우 이질적인 개인의 복합적인 의사결정인 데다가(Gauthier, 2007) 실증적으로는 제한된 관측치와 내생성 문제 등으로 사회과학적인 분석이 어려운 것이 사실이다.

본 연구는 하버드 대학의 Raj Chetty 교수가 이끄는 연구팀의 기회불평등분석 방법론과, 2014년부터 공개되기 시작한 개인수준의 국민건강보험 빅데이터를 이용하여 환경과 출산율과의 관계에 대한 해답의 단초를 제공하는 것에 그 목적이 있다. 이를 위해 지난 10여 년 동안 약 259만여 출산경험자의 건강보험 기록을 통해 그들의 출산 연도와 거주지 주소 등을 포함하는 광범위한 DB를 구축하였다. 그리고 기존의 TFR 등이 환경에 의한 이주자를 포함하기 때문에 지역특수적인 출산환경의 효과를 정확히 나타낼 수 없는 것을 극복하기 위해 그 기간 한 곳에 계속 거주한 여성의 출산기록을 바탕으로 출산율의 지역적 차이를 다시 계산한다. 또한, 계산된 지수를 출산에 유리한 환경을 나타내는 지수로 사용해 이러한 환경에 이주한 약 80만 여성들의 출산기록이 환경에 따라 차별적인 특징을 지니는지 분석하였다. 출산율과 환경의 관계에 대한 준실험적 연구방법(Quasi-Experimental Method)이라고 할 수 있는 본 연구는 이후 보다 구체적인 정책적 함의 도출이 가능한 연구를 수행하기 위한, 예비적 분석으로서의 의미가 있다.

실증결과, 본 연구에서 계산한 시군구별 출산환경은 지역마다 차이가 있었고 이러한 환경에 이주한 이주자의 출산율에 유의미한 영향을 미치고 있음을 확인할 수 있었다. 구체적으로, 출산에 좋은 환경이 초기에 출산율을 상승시키는 부

구의 지역유입을 촉진할 수도 있음을 고려할 때 정책의 효과에 대한 실증분석 결과가 과장될 수 있음을 지적하고 있다.

2) 「출산율 1위 '해남의 기적'은 끝났나? 출산 장려금만 먹튀 수두룩」, 조선일보, 2021년 1월 24일 자 기사.

분 중 약 65% 정도는 선별효과에 의한, 즉 출산 가능성이 높은 세대를 모으는 효과에 의한 것이고 실제로 출산율을 상승시키는 효과는 그 이후에 서서히 나타나 6년 정도에 가장 정점에 도달했다가 이후 정체된다고 볼 수 있다. 이는 지속가능한 출산정책만이 시차를 두고 그 효과가 있을 수 있다는 것을 의미한다. 제2장은 연구의 배경을, 제3장은 연구 디자인에 대한 소개와 실증결과를 논의한다. 제4장에서 정책적 함의와 본 연구의 한계를 언급하고 맺을 것이다.

Ⅱ. 연구의 배경

출산율 결정요인과 저출산정책의 효과에 관해 국내외 많은 연구가 존재한다. 예를 들어, 1970년도부터 1990년도까지 OECD 국가별 시계열 자료를 분석한 Gauthier and Hatzius(1997)의 연구는 평균보다 25% 높은 가족수당이 출생아 수를 0.07 정도 상승시켰다는 결과를 제시하였다.

Milligan(2005)은 1988~1997년까지 캐나다 퀘벡주에서의 저출산정책 효과에 대한 실증분석을 통해 정책의 효과는 평균 12%, 최대 25% 정도의 출산율 증가라고 주장하였다. 호주의 경우에도 자녀수당(Baby Bonus)의 지급과 출산율이 양의 상관관계를 보였다는 연구결과가 있다(Drago et al., 2011). González (2013)의 연구는 2007년 스페인에 도입된 아동수당이 부분적으로는 낙태율의 감소를 통해 출산율을 유의하게 증가시켰으며 여성의 노동공급 또한 감소시키는 효과를 낳았다고 주장하였다.

한국의 연구로는 예를 들어 이삼식 외(2011)가 OECD 10개국의 패널자료를 이용해 조혼인율, 초산연령, 여성경제활동참가율, 일인당 국민소득, 남성 대비 여성 대학진학 비율, 양성평등지수, GDP 대비 보건정책 지출비율 등 다양한 사회경제변수들을 출산율 결정요인으로 추출하고 이를 이용해 출산율 예측모형의 개발을 시도한 바 있다. 박창우 · 송헌재(2014)는 2005~2010년 전국 230개 지방자치단체를 대상으로 출산장려금정책의 효과를 고정효과모형으로 분석하여 출산장려금이 출생아 수를 첫째는 0.4%, 둘째는 0.44% 증가시키나 셋째는 유의하게 증가시키지 않는 것으로 추정하였다.

이상협 외(2016)는 보육료 지원 또는 양육수당 지원의 두 정책을 동시에 고려하고 패널 GMM 등의 계량경제학의 방법론을 이용해 시군구별 출산장려금, 아동인구 대비 보육시설 수 등의 환경이 유배우 출산율과 무배우 혼인율에 상반된 효과를 보인다고 주장하며 지난 10년 동안의 출산장려정책이 실패했다는 다수의 견해에 의문을 제기하였다. 조영태 외(2018)는 시군구별로 2008년부터 2016년까지 인구, 주택, 고용 등 다양한 사회경제적 요인들을 출산율과 회귀분석해 출산율 결정요인을 선별하고 이를 이용해 한국의 출산예측모형 개발을 시도하였다. 이밖에도 김민영・황진영(2016)이나 송헌재・우석진(2015) 등이 주택가격이나 보육지원정책이 출산율에 미친 영향을 실증분석하여 대체적으로 출산에 유(불)리한 환경이 출산율에 긍정적(부정적)인 영향을 주고 있음을 주장하였다.

사실 이 연구들은 다양한 데이터와 실증기법을 이용해 내생성이나 역인과관계 등의 문제 극복을 시도하며 출산을 결정하거나 영향을 주는 요인들을 분석하고 있지만 한계 또한 지닌다. 예를 들어 이삼식 외(2011)가 주장하듯이 출산에는 교육, 노동, 주택, 보건의료, 보육, 가치관, 문화 등의 복합적인 요인들이 작동하기 때문에 이를 모두 고려해 실증분석을 하기는 어려울 뿐만 아니라, 고려할 수 있는 다양한 변수들을 동시에 출산율 결정요인으로 고려한다고 하더라도 이 연구들의 기본적인 문제점은 출산율을 결정하는 요인 중 장소효과(Place Effect)에 해당하는 부분과 선별효과(Selection effect)라 할 수 있는 부분을 구분하지 못하는 것이다.

장소효과는 어떤 정책이나 요인의 지역적 변화가 그 지역에 사는 사람들의 출산율에 영향을 주는 효과라 할 수 있고 선별효과는 상이한 특성을 지닌 사람들의 이주에 의한 효과이다. 현실적으로 관찰 가능한 데이터가 과거의 지역별 정책과 지역 거주민의 출산율 등이니 이들을 회귀분석하는 과정에서 해당 기간에 환경이 좋은 곳으로 이주한 사람들에 의한 선별효과를 구분하지 못하는 것이다. 예를 들어 주거에 대한 지역특수적인 정책이 지역의 출산율을 상승시킬 수도 있지만 낮은 주거비용이 출산 가능성이 높은 특성을 지닌 개인들을 그 지역으로 모을 수도 있다. 또한 출산에 유리한 환경에 이주한 개인들이 환경에 영향을 받아 출산율이 상승하는 노출효과(Exposure Effect)도 정의할 수 있다.

이러한 효과들이 뒤섞이면 정책의 효과에 대한 실증적인 결론들은 과장될 수 있다.

만약 장소효과나 노출효과가 존재한다면 지역특수적인 정책의 효과가 존재한다는 것이고 이러한 장소에 사람들이 살게 하거나 이러한 장소의 특징들을 식별해 유사한 환경을 만든다면 출산율을 상승시키는 데 도움이 될 것이다. 반면에 장소효과 혹은 노출효과가 미미하다면 일부 지자체의 출산장려금 같은 지역특수적인 정책은 재고되어야 할 것이다.

그렇다면 이러한 효과들을 어떻게 구분할 것인가? 본 연구는 효과들을 식별하기 위해 다음과 같은 질문에 대한 답을 구한다. 즉 "출산환경이 양호한 환경으로 이주한 여성의 출산 가능성은 증가하는가?" 이 질문에 대한 실증적 답이 긍정적이라면 장소효과가 존재한다는 것을 의미할 것이다.

그렇다면 이것을 어떻게 실증할 수 있을까? 아마도 이것에 대한 이상적인 실증 환경은 요즘 주목받는 무작위 실험일 것이다. 즉 선택 편의가 제거되도록 개인특성이 상이한 사람을 무작위로 선정, 집단을 만들고 이들을 상이한 환경으로 이주시킨 뒤 10년 후 두 집단의 출산율을 비교하는 것이다. 물론 이러한 실험은 현실에서 가능하지 않다. 대안적으로 과거의 데이터를 이용해 실험과 유사한 효과를 내는 준실험적 방법을 생각해볼 수 있는데 일단 충분한 관측치가 확보되느냐의 문제는 차치하고서라도 출산은 내생적인 선택이기 때문에 이사 간 사람들의 출산율에 대한 단순 비교는 장소효과와 선별효과가 섞이게 마련이다. 사실 준실험적 방법은 결과에 영향을 미치지만 우리가 통제하고 싶은 요소들을 데이터를 통해 식별해낼 필요 없이 필요한 요소의 효과만을 볼 수 있다는 데 그 장점이 있다.(Chetty and Hendren, 2018a) 다음 절에서 출산율과 환경의 관계를 살펴보기 위해 이 방법을 어떻게 디자인할 수 있는지 살펴보자.

Ⅲ. 실증작업

본 연구의 디자인은 우리가 흔히 접하는, 비교적 소수를 대상으로 한 서베이 자료를 이용해서는 실행 가능하지 않다. 후술하겠지만 본 연구의 연구 디자인

이 출산여성 중 해당 기간에 계속해서 거주한 여성과 이주한 여성을 구분하고 이들의 10여 년간의 출산기록을 바탕으로 하고 있기 때문이다. 우리가 다루는 건강보험 빅데이터는 전수조사에 가까운 관측치를 포괄하기 때문에 계량경제학적으로 몇몇 가정들을 잘 배치하고 연구 디자인을 잘 설계한다면 이러한 요소들을 통제하고 우리가 알고 싶은 과학적인 사실들을 유추할 수 있다.

이 연구의 방법론은 하버드 대학의 Raj Chetty 교수가 이끄는 연구팀의 기회불평등분석 방법론에서 영감을 받은 바 크다. Chetty et al.(2014, 2016), Chetty and Hendren(2018a, 2018b) 등은 American Dream이 1980년대 이후 사라졌다는 사실을 사회과학적으로 밝혀내기 위해 세금자료와 연말정산 등의 빅데이터를 이용, 기회불평등을 세대 간 소득이동성, 예를 들어 하위 25% 소득의 가정에서 자란 자녀세대의 평균적 소득분위 등으로 계산하고 이를 가입자의 거주 시군구별로 분류하여 전국적인 기회불평등 지도를 구축하였다.

또한 이러한 지역 간 기회불평등 차이의 원인을 ① 환경의 인과적 영향(장소효과) ② 인구구성이나 자산 등 각 지역 거주민들의 체계적 차이(선별효과)로 나누고 이 둘을 구분하기 위한 실증분석을 시도한다. 사실 이에 대한 가장 이상적인 실증방법은 앞 절에서 언급하였듯이 무작위로 대상을 선정하여 다른 환경에 노출시키고 수십 년 후 그 결과를 분석하는 실험적 방법일 것이다. 하지만 이 방법이 가능하지 않기 때문에[3] 그들은 준실험적 방법을 도입했다. 여기에서 준실험적 방법은 “영구거주민 간 더 높은 이동성을 보이는 지역에 이사 간 아이들이 더 높은 성과를 보이는가?”라는 질문에 답하기 위해 이주자의 성과와 이주한 지역의 기회불평등 차이를 회귀분석하는 것을 의미한다.

본 연구에서도 영구거주자의 성과(출산) 차이를 시군구별로 구분하여 이들 지역에 이주한 여성이 더 높은 성과(출산율)를 보이는가를 분석함으로써 환경이 출산에 미치는 영향을 탐구해본다.

---

3) 예외적으로 사회과학적인 실험이 가능했던 경우가 취약계층에 주택바우처를 제공했던 미국의 Moving to Opportunity 실험이다. 이와 관련된 논의는 Katz et al.(2001), Oreopoulos (2003), Kling et al.(2007), Ludwig et al.(2013), Chetty et al.(2016), Bergman et al. (2019) 등 참고.

1. 연구설계

먼저 환경의 효과를 측정하기 위해 앞서 언급한 Raj Chetty교수의 연구와 유사하게 이주자가 직면한 환경을 나타내기 위해 그곳에 이미 거주하고 있는 거주자를 이용한 출산환경지수를 산출한다. 본 연구에서는 대상 기간에 하나의 시군구에서 계속 거주한 거주자(여성)를 영구거주자로 정의하여[4] 이 영구거주자만을 대상으로 한 시군구 c에서의 평균 출생아 수 y_c를 식(1)과 같이 계산해 이용한다.

$$y_c = \frac{\text{영구거주자의 출생아 수}}{\text{영구거주자 수}} \quad (1)$$

이런 방식으로 234개 시군구에 대한 영구거주자의 평균 출생아 수를 계산해 출산에 유리한 환경의 정도를 측정하는 대리지표로 사용한다.[5] 그리고 우리는 노출효과 즉, 출산에 유리한 환경에 이주한 이주자는 이러한 환경에의 노출로 인해 출산할 가능성이 상승한다고 가정한다.

그리고 노출효과는 다음과 같이 측정한다. 먼저 해당 기간 거주지를 이전한 여성을 대상으로 이주자의 출생아 수 y_{id}를 다음과 같이 계산한다.

$$y_{id} = \alpha_m + b_m y_{cd} + \theta_i \quad (2)$$

여기에서 m은 이주 후 좋은 환경에 노출된 연수이고 α_m은 상수항, b_m은 좋은 환경이 출생아 수 증가에 미치는 영향을 나타내는 계수, y_{cd}는 이주자가 이주한 이주지(Destination)의 기대 출생아 수 y_c이고 이 y_c는 앞에서 설명하였듯이 234개 시군구 영구거주자의 평균 출생아 수이다. θ_i는 개인특성이나 개인의 출산을 결정하는 다른 요소이다. 본 연구에서는 출산을 결정하는 개인특성은 고정되어 있고 앞서 언급하였듯이 여성이 출산에 유리한 환경으로 이주하면

4) Raj Chetty의 연구에서는 이들이 한 곳에 계속 거주한 기간이 분석대상기간에만 국한된 것임에도 불구하고 이들을 Permanent Resident, 즉 영구거주자로 정의하였다. 이는 환경의 효과를 분석하기 위한 (조작적) 정의일 뿐이다. 본 논문에서도 이를 따른다.

5) 이러한 대리지표로 시군구별 TFR을 사용하지 못하는 이유는 앞서 언급하였듯이 TFR이 장소효과뿐만 아니라 선별효과까지 모두 포함하고 있기 때문이다.

그러한 환경에 영향을 받아 출생아 수는 증가한다고 가정한다.

식 (2)는 Chetty et al.(2014, 2016), Chetty and Hendren (2018a, 2018b) 등이 기회불평등의 환경영향을 측정하기 위해 고안한 회귀식과 동일한 형태로서 다만 출산율 연구의 맥락에서 응용했다고 볼 수 있다. 기회불평등과 출산은 무수히 많은 요소에 의해 결정되기 때문에 실증분석의 어려움이 있다는 공통점이 있지만 이 식은 개인특성과 환경의 영향 중 출산에 대한 환경의 영향을 측정하기 위해 고안된 것이다. 즉 "출산환경이 양호한 환경으로 이주한 여성의 출산 가능성은 증가하는가?"에 대한 질문에 답하기 위해 준실험적 방법을 사용, 즉 실험이 불가능한 상황에서 과거의 데이터를 이용해 마치 실험을 하는 것과 같은 효과를 기대하는 것이다.

여기서 중요한 것은 출산을 결정하는 매우 다양한 개인특성들인 θ_i를 관찰할 수 없는 상황에서 위와 같은 방식으로 노출효과가 측정되려면 θ_i가 y_{cd} 와 상관되는 내생성 문제가 존재하지 않아야 한다는 것이다. 이를 충족시키는 세 가지 방법이 있으며 첫째가 분석대상 선정 시 보고자 하는 요소를 제외한 다른 요소들은 완벽하게 동일한 집단을 대상으로 관찰하는 것인데 이는 현실적으로 불가능하다. 둘째는 무작위로 대상을 선정해서 이질적인 요소가 서로 상쇄되게 하는 실험방식을 채택하는 것인데, 예를 들어 무작위로 이주자와 이주지를 선정해서 수년 후 그 결과를 관찰할 경우 개인의 특성 θ_i는 y_{cd}와 상관되지 않고 따라서 b_m은 정확한 노출효과가 될 것이다.

하지만 무작위 실험이 불가능한 상황에서 현실의 데이터로 이를 회귀분석하면 관찰할 수 없는 θ_i 때문에 b_m은 환경의 영향인 β_m을 포함하여 다음과 같이 측정된다.

$$b_m = \beta_m + \delta_m \tag{3}$$

여기에서 δ_m은 $\frac{cov(\theta_i, y_{cd})}{var(y_{cd})}$이다. 이는 y_{cd}가 θ_i가 상관되어 y_{id}에 영향을 주는 경우, 예를 들어 좋은 환경으로 이주한 가정일수록 더 출산 가능성이 높은 개인특성을 지니고 있어 기대 출생아 수가 많은 경우이다. 이 항이 앞서 언급한 선별효과를 나타내는 부분이라고 할 수 있다.

본 연구에서는 이러한 내생성 문제를 방지하기 위해 Raj Chetty의 연구와 동일한 가정을 도입한다. 그것은 출산을 결정하는 개인의 특성 등이 이주장소에 따라 다를 수는 있지만 이주 연도나 이주의 시기와는 무관하다는 가정이다. 즉,

$$\delta_m = \delta \text{ for all m} \tag{4}$$

따라서 (2)식으로 구한 b_m은 노출효과와 선별효과가 섞여 있지만 환경노출 연수 m과 m-1을 차감해주면 δ가 소멸되어 1년의 노출효과만을 얻을 수 있다. 1년의 노출효과(γ)는 좋은 환경으로 이주했을 경우 γ에 비례해서 기대 출생아 수가 증가한다는 것을 의미한다.

$$\gamma = b_m - b_{m-1} = \beta_m - \beta_{m-1} (\text{if } \delta_m = \delta \text{ for all m }) \tag{5}$$

출산을 결정하는 개인의 특성 등이 이주의 시기와는 무관하다는 가정은 다소 무리한 것일 수도 있고 따라서 추가적인 검증이 필요하다. 예를 들어 이주시기와 출산시기의 차이가 많이 나는 가정이 원래 출산을 더 많이 혹은 적게 하는 가정일 수 있다. 이에 대해서는 기본적인 분석을 수행한 후 추가적인 분석을 통해 논의할 것이다.

2. 데이터 구축

본 연구는 국민건강보험공단이 제공하는 국민건강보험 빅데이터 DB 중 맞춤형 DB를 이용한다. 이 DB는 국민건강보험에 가입한 한국민들의 성별, 보험료, 건강보험 가입형태 등의 기본적인 자격데이터를 포함하여 각종 진료, 처방 데이터를 포함하는 방대한 DB를 연구자가 원하는 형태로 가공해서 구축한 것이다. 중요한 것은 이 데이터가 비식별 처리가 되어 있는 개인별 수준의 미시데이터라는 점이다. 개인별 수준의 미시데이터는 계량분석을 할 때 유리한 환경을 제공한다.

본 연구의 연구방법론에 맞추어 구축된 데이터는 2010년부터 2018년까지 출산한 약 259만 출산여성의 거주지 등의 정보를 포함하고 있는 기본적인 자격 및 보험료 데이터셋이다. 이 데이터셋은 출산관련 행위코드[6]를 진료내역(T30)

에서 검색하여 출산자를 식별하고 이를 다시 명세서(T20) 및 자격 테이블과 매칭하는 방식으로 구축하였다.

<표 1>, <표 2>는 이렇게 구축한 데이터셋의 기초통계량을 나타내는 표들이다. 먼저 <표 1>은 전체 데이터셋을 2010년부터 2018년까지 여성들의 이주 횟수별로 나눈 관측치의 수이고 <표 2>는 전체 관측치와 이주 횟수가 0과 1인 관측치만 따로 구분하여 그 특성을 나타낸 표이다. 전체 259만여 명의 출산여성 중 해당 기간 시군구 단위 기준으로 한 번도 이주하지 않은 건은 약 79만여 건이고 한 번만 이주한 경우는 약 83만여 건이다. 영구거주자에 대한 분석은 이들 79만여 건의 관측치를 대상으로 하였고 노출효과는 분석 편의상 여러 번 이주한 경우를 제외하고 한 번만 이주한 경우인 83만여 건을 대상으로 하였다.

〈표 1〉 이주 횟수

(단위 : 건, %)

이주 횟수	관측치	비율	누적비율
0	795,782	30.71	30.71
1	830,968	32.06	62.77
2	581,537	22.44	85.21
3	263,140	10.15	95.36
4	91,070	3.51	98.87
5	23,816	0.92	99.79
6	4,696	0.18	99.97
7	645	0.02	100
8	35	0	100
전 체	2,591,689	100	

자료 : 국민건강보험공단 맞춤형 DB.

6) 출산관련 행위코드는 다음과 같은 정상분만과 제왕절개분만의 행위코드를 모두 포함하였다. 정상분만 : R3131, R3133, R3136, R3138, R3141, R3143, R3146, R3148, R4351, R4353, R4356, R4358, R4361~2, R4380, RA311~8, RA361~2, RA380, RA431~4. 제왕절개분만 : R4507~10, R4514, R4516~20, R5001~2.

〈표 2〉 데이터의 특성

(단위 : 건, 세, 분위, 회, 명)

변수명	관측치	평균	표준편차	최소값	최대값
전 체					
나이	2,591,689	28.570	4.743	7	49
소득(20분위)	2,458,272	10.367	5.207	1	20
이주 횟수	2,591,689	1.273	1.175	0	8
출생아 수	2,591,689	1.461	0.634	1	6
이주 횟수=0					
나이	795,782	29.600	4.940	7	49
소득(20분위)	756,261	10.535	5.146123	1	20
출생아 수	795,782	1.325	0.528	1	6
이주 횟수=1					
나이	830,943	28.478	4.698	7	49
소득(20분위)	792,183	10.321	5.204	1	20
출생아 수	830,943	1.387	0.553	1	6
출생아 수(이주 후)	830,943	0.807	0.720	0	5

자료 : 국민건강보험공단 맞춤형 DB.

전체 데이터셋은 그 관측치가 2010년부터 2018년까지의 모든 출산여성의 수 259만여 건이고 이들의 첫째부터 여섯째까지 출산 연월과 주소지 변동 등을 포함하고 있다.[7] 출산여성의 평균 나이는 2010년 기준 28.570세[8]이고 모두 20분위로 나누어져 있는 평균 소득분위는 10.367 분위이며 이들의 평균 이주 횟수(시군구별)는 1.273회, 평균 출생아 수는 1.461명이다. 해당 기간에 전혀 이주하지 않은 경우(이주 횟수=0)인 영구거주자 데이터셋은 약 79만 여 건으로 이들의 평균 나이는 29.600세, 평균 소득분위는 10.535분위, 평균 출생아 수는 1.325명이다. 또한 해당 기간 1번만 이주한 경우(이주 횟수=1), 즉 이주자 데이터셋은 약 83만 여 건으로 이들의 평균 나이는 28.478세, 평균 소득분위는 10.321분위, 이주 후 평균 출생아 수는 0.807명이다.[9]

7) 본 연구의 데이터셋은 약 19만 7천여 건의 쌍둥이 출산을 포함하고 있으나 분석 시 이를 모두 한 명의 출산으로 대체하였다. 그 이유는 비록 쌍둥이 출산이 다음 아이 출산에 영향을 주기는 하지만 기본적으로 그것이 출산을 결정하는 개인의 특성, 혹은 출산에 유리하거나 불리한 환경 등에 의해 영향을 받는 요소가 아니기 때문이다. 또한 7번째 아이 이상의 출산도 존재하나 관측치가 적어 본 연구에서는 제외하였다.

8) 나이의 경우 1919년생 등의 이상치가 전체 259만여 건 중에 약 100여 건 존재한다. 이러한 이상치는 모두 제외하고 해당 기간(2010~2018년) 가임 여부를 고려하여 2010년 기준 7세부터 49세까지로 제한하였다.

3. 실증결과

가. 예비적 분석

먼저 예비적 실증분석으로서 해당 기간 출산경험 여성으로 이루어진 영구거주자 데이터셋을 이용해 계산한 시군구별 평균 출생아 수(y_c)를 살펴보자. y_c는 2010년부터 2018년까지 동일한 시군구의 영구거주자 개인별로 진료 및 상병 테이블에서 분만시술코드를 검색해 출산 연도를 추출하고 이때의 거주 시군구(코드 5자리)를 기록한 후 이를 다시 시군구별로 재계산해서 산출하였다. <표 3>은 y_c의 기초통계량이다.

〈표 3〉 y_c의 기초통계량

(단위 : 명)

변수명	관측치	평균	표준편차	최소값	최대값
y_c	239	1.317	0.059	1.182	1.470

자료 : 국민건강보험공단 맞춤형 DB.

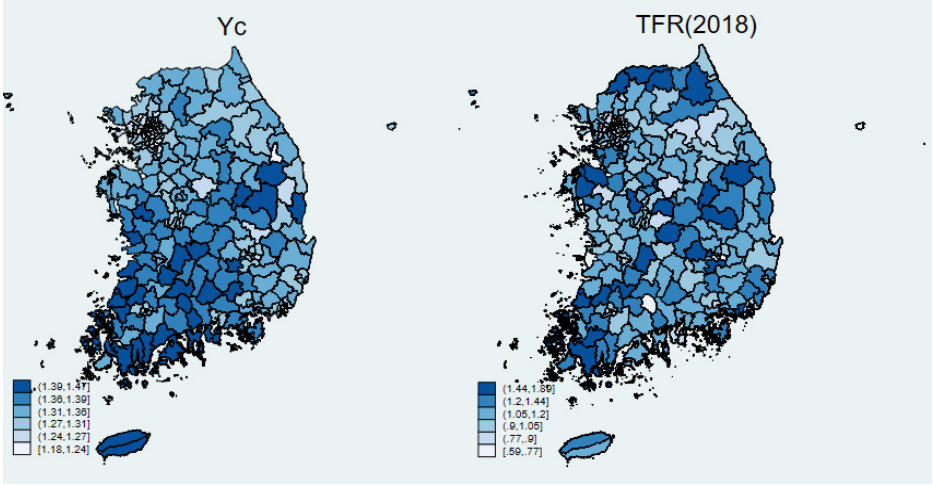
모두 239개의 시군구에 대해서 영구거주자의 출산기록에 따라 y_c가 계산되었고 그 평균은 1.317, 표준편차는 0.059이다. 해당 기간, 시군구별로 평균 1.317명의 아이를 출산한 것이다.

그렇다면 이렇게 계산한 y_c는 시군구별로 어떤 패턴을 보일까. [그림 1]은 전국의 시군구별 y_c를 표시한 것이고 [그림 2]는 서울의 구별 y_c를 나타낸 것이다. 특히 두 그림 모두 TFR과 비교하여 나타내었는데 두 지표 간 상이한 패턴을 알 수 있다.[10]

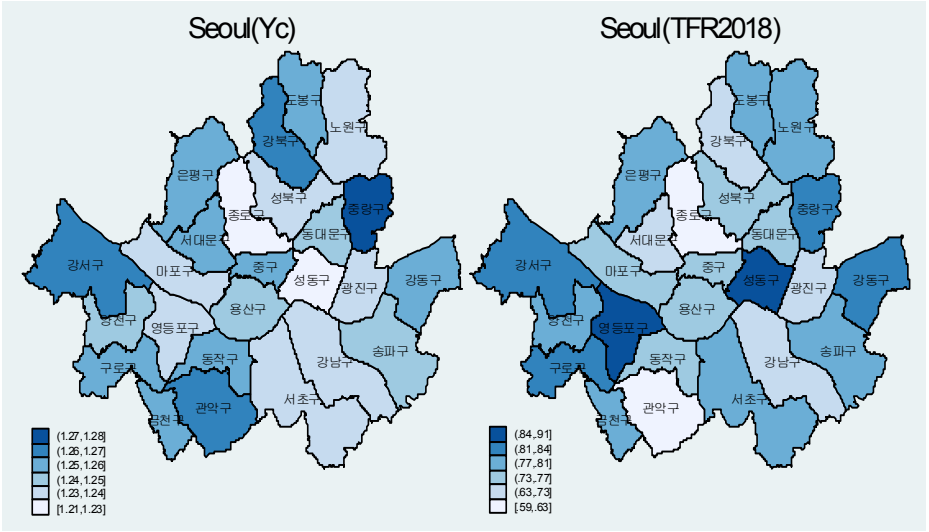
9) 한편 건강보험의 가입형태(Type)는 모두 6가지로 구분되어 있는데 직장가입자 세대주와 세대원, 지역가입자 세대주와 세대원, 의료급여 세대주와 의료급여 세대원 등이다. 전체 데이터셋에는 직장가입자(세대원 포함)와 지역가입자(세대원 포함), 의료급여세대(세대원 포함) 수가 각각 약 190만 명, 66만 명, 3만 명이었고 영구거주자 데이터셋은 각각 58만 명, 20만 명, 1만 명, 이주자 데이터셋은 각각 61만 명, 21만 명, 1만 명이다.

10) 사실 이 차이는 기본적으로 두 지표의 산정방식이 다르기 때문에 발생하는 것이다. 특히 지역의 TFR(2018)은 2018년에 그 지역에 거주하는 15세에서 49세까지 모든 여성의 수

〔그림 1〕 전국의 시군구별 영구거주자 y_c와 합계 출산율(TFR, 2018)



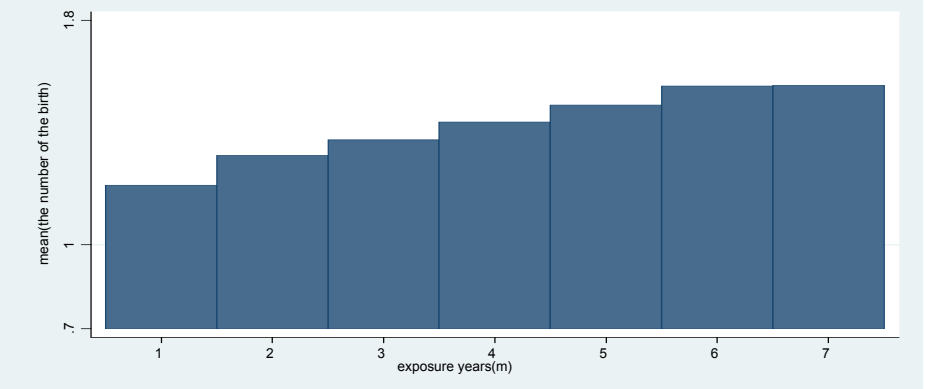
〔그림 2〕 서울의 구별 영구거주자 y_c와 합계 출산율(TFR, 2018)



가 분모에 고려되지만 y_c는 2010년부터 2018년까지 출산 경험자만을 대상으로 한다는 것, 그리고 각 시군구별로 상이한 정책 등이 고려되어야 하기 때문에 지표 차이의 원인에 대한 보다 엄밀한 분석이 요구된다. 하지만 영구거주자의 지표와 다르게 TFR에 인구이동의 효과가 포함된다는 것은 흥미로운 사실이다. 이를 반영하여, 예를 들어 강원도 태백시나 서울 성동구는 TFR은 높지만 y_c는 매우 낮다. 두 지표의 지역별 차이에 관한 논의는 김준일(2021) 참조.

그렇다면 이러한 환경에 이주한 이주자의 출산율은 어떤 영향을 받았을까? 이제 노출효과에 대한 분석결과를 살펴보자. 회귀분석에 앞서 예비적 분석을 수행하였다. 여성이 출산에 유리한 환경으로 이주하면 그러한 환경의 영향으로 출생아 수는 증가한다고 가정한 노출효과에 관한 분석결과를 [그림 3]을 통해 살펴보자. 이주자 데이터셋에서 이주자들의 출발지(Origin) 및 이주지(Destination)에 시군구별 평균 출생아 수(y_c)와 출생아 수를 매치시켜 데이터셋을 구성하였다. 그리고 y_c 기준 하위 10%의 출발지에서 상위 10%[11] 이주지로 이주한 관측치(극단적 이주자 : Extreme Movers)만을 남겨 그 특성을 살펴보았다. [그림 3]에서 보는 것과 같이 이주 후 좋은 환경에의 노출 연수가 증가할수록 이들의 해당 기간(2010~2018년) 평균 출생아 수는 증가하였다. 예를 들어 이주 후 좋은 환경에의 노출기간이 1년밖에 되지 않는 여성의 평균 출생아 수는 1.21명인데 7년 노출된 경우는 1.57명이었다. 좋은 환경에의 노출기간이 길었기 때문에 평균적으로 더 많은 아이를 출산한 것인가, 아니면 출산을 할 가능성이 높은 인구가 좋은 환경으로 이주한 결과인가? 즉 이것은 장소효과의 결과인가, 아니면 선별효과의 결과인가? 이 기술통계값만을 가지고 판단하기는 어렵다. 우리의 연구 디자인을 이용해 좀 더 구체적으로 분석할 필요가 있다.

〔그림 3〕 극단적 이주자의 평균 출생아 수



주 : y축은 평균 출생아 수, x축은 노출 연수.

11) 상·하위 10%의 기준은 이주자의 특성을 보여주기 위해 임의로 선택한 것이다. 다른 기준으로 구성해도 그 결과는 대동소이했다.

나. 기본적 회귀분석

이제 본 연구의 주요한 연구 디자인이라고 할 수 있는 식 (2)에 대한 실증분석을 이용해 노출효과를 계산해보자. Ⅲ.1에서 논의한 것처럼 b_m은 노출효과인 β_m과 선별효과인 δ_m을 포함한다. 만약 $\delta_m = \delta$ (for all m)이라면 1년의 노출효과(γ)을 구할 수 있다. <표 4>는 m의 값에 따라 각각 (1)식을 회귀분석해서 그 결과를 모아 놓은 것이다. 예를 들어 m=1은 2017년 이주해서 이주지 환경에 노출 연수가 1년인 여성을 대상으로 그들의 이주지 y_c와 이주 후 출생아 수를 회귀분석한 것이다. 그 계수값인 b_m의 값은 m=1일 때 0.030이었다가 m=2에서는 0.464로 증가하였다. 환경에의 노출 연수가 3년, 4년으로 증가함에 따라 b_m의 값도 증가하여 m=6일 때 1.488로 최대가 되고 이후 정체한다. b_m의 값이 이주자의 환경에 대한 노출 연수가 증가할수록 비례해서 거의 선형으로 증가하는 것이다. 이는 노출효과가 존재함을 의미한다.

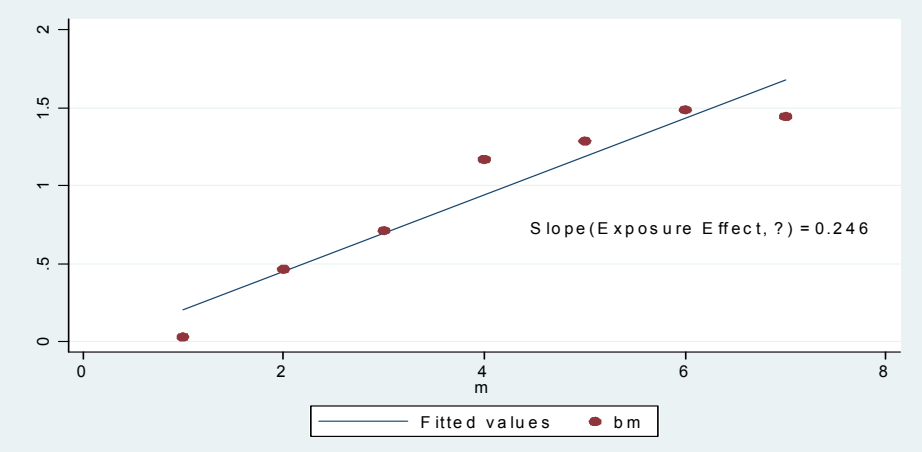
〈표 4〉 회귀분석결과

Variable	m=1	m=2	m=3	m=4	m=5	m=6	m=7
b_m	0.030	0.464	0.712	1.169	1.287	1.488	1.446
	(0.041)	(0.045)	(0.049)	(0.051)	(0.051)	(0.047)	(0.042)
_cons	0.439	0.001	-0.206	-0.686	-0.740	-0.906	-0.700
	(0.054)	(0.059)	(0.064)	(0.068)	(0.068)	(0.062)	(0.055)
N	80118	87087	90618	93415	97742	116145	124643

주: ( ) 수치는 표준오차임, _cons는 상수항, N은 관측치.

노출 연수 m과 b_m의 선형관계를 [그림 4]를 통해 살펴볼 수 있다. 그 기울기는 0.246으로 우리의 모델에서 노출효과에 해당한다. 즉 1년 더 환경에 노출될수록 좋은 환경에 의한 출생아 수는 평균적으로 0.246씩 증가한다. 예를 들어 y_c가 1만큼 더 높은 환경에 이주하면 1년에 0.246명씩 기대 출생아 수가 증가하는 것이다.

〔그림 4〕 노출효과



주: Fitted values는 적합값, bm은 회귀계수, y축은 회귀계수의 크기, x축(m)은 노출연수, Slope(exposure effect, γ)는 노출효과를 나타내는 기울기.

[그림 4]의 결과는 Chetty and Hendren(2018a)의 기회불평등 분석결과 그림과 유사하다. 그들의 연구에서 영구거주자의 성과와 이주자 자녀의 성과는 선형관계, 즉 영구거주자의 성과가 좋은 곳에 이주한 이주자 자녀의 성과는 좋은 환경에 노출된 연수가 증가할수록 영구거주자의 성과값으로 수렴하였다.

그렇다면 선별효과는 얼마나 될까? 우리가 구한 b_m에는 노출효과인 β_m과 선별효과 δ_m이 어떤 비율로 결합되어 있는지 알 수 없다. Chetty and Hendren (2018a)에서는 환경이 영향을 줄 수 없는 성인 이후 이주자의 계수로써 추론하였다. 우리의 맥락에서는 다음과 같다. 논리적으로 이주 후 1년이나 2년 차에는 환경에의 노출기간이 짧고 임신 결심과 임신, 그리고 출산까지 최소 10개월 이상 걸린다는 것을 고려하면 노출효과는 매우 적을 것으로 추론할 수 있다. 즉 b_1과 b_2에는 새롭게 이주한 환경에의 영향은 존재하지 않고 대부분 이전 거주지의 환경과 개인특성에 의한 선별효과가 포함되어 있다는 것이다. 그렇다면 예를 들어 3년 차(m=3)의 효과 0.712 중 약 65%(0.464/0.712*100)는 선별효과에 의한 것이라고 볼 수 있다.[12)]

12) 물론 이는 대략적인 계산과 추론이다. 빅데이터에는 이주시기가 월 단위로 되어있어 좀 더 엄밀한 분석도 가능하였지만 그렇게 한다면 데이터 분량이 방대하여 분석이 너무 복잡

이상의 결과를 종합해보면 출산에 좋은 환경이 초기에 출산율을 상승시키는 부분 중 상당 부분은 선별효과에 의한, 즉 출산 가능성이 높은 세대를 모으는 효과에 의한 것이고 실제로 출산율을 상승시키는 효과는 그 이후에 서서히 나타나 6년 정도에 가장 정점에 도달했다가 이후 정체된다고 볼 수 있다.

다. 추가적 회귀분석

본 연구의 핵심 가정은 앞서 언급하였듯이 출산을 결정하는 개인의 특성 등이 이주장소에 따라 다를 수는 있지만 이주 연도나 시기와는 무관하다는 가정이다. 즉 y_{cd}가 높은 곳에 이주하는 세대의 경우 θ_i가 높아 출생아 수 y_{id}도 높지만 이 선별효과는 $\delta_m = \delta$(for all m)를 가정함으로써 1년 노출효과를 계산할 때 서로 소거된다는 것이다. 이 가정이 성립하지 않을 경우 추정치에 편의가 존재할 가능성이 있다. 이에 대해 자세히 논의해보자.

우선 세대 내 출산을 결정하는 요인인 θ_i를, 시간에 따라 변화하지 않는 $\overline{\theta_i}$와 변하는 $\tilde{\theta}_i$, 두 부분으로 나누어 생각해보자.

$\overline{\theta_i}$를 우선 고려했을 때, 가정이 성립하지 않을 경우는 예를 들어 $\overline{\theta_i}$가 높은 세대일수록 좋은 환경으로 일찍 혹은 늦게 이주할 경우 즉, $\delta_m \neq \delta_{m-1}$이다. 우리의 데이터에서는 2010년대 초반 혹은 후반에 집중적으로 출산 가능성이 높은 세대가 좋은 환경으로 이주했을 경우라고 볼 수 있다. 하지만 체계적으로 이런 현상이 발생한다고 생각할 타당한 근거를 찾기 힘들다. 정부정책 중 세종시나 혁신도시 건설이 가능한 경우인데 본 연구의 데이터에는 이들 도시가 2010년대 중간에 나타나기 때문에 영구거주자를 선별할 수 없어 누락되어 있다.[13]

$\delta_m \neq \delta_{m-1}$인 또 하나의 경우는 연령이다. 출산이 생애주기에 따라 결정되는 의사결정이다 보니 연령의 영향을 많이 받는다. 만약 특정 연령대에서 이주와 출산이 이루어진다면 그리고 $\overline{\theta_i}$가 높은 세대가 집중적으로 특정 연도에 이주를

해지는 한계가 있었다.

13) 혁신도시는 2012년에 부지조성이 완료되나 실제 이전완료시기는 지역별로 차이가 있지만 2017 · 2018년도 즈음이다. 세종시는 2007년 착공해서 2012년 첫 입주를 하였다(혁신도시 홈페이지 : http://innocity.molit.go.kr/v2/, 2021년 3월 1일 검색).

많이 하게 되면 $\delta_m \neq \delta_{m-1}$일 수 있다. 데이터셋에서 실제로 연령과 이주시기의 특성을 살펴보면 2010년대 초반에 이사한 여성의 경우 후반에 이사한 여성보다 이주 당시 연령이 더 어린 것을 알 수 있었다. 이런 경우 연령이 어리므로 출산 가능성이 높고 이것이 어떤 체계적인 차이를 가져올 수 있을 것이다. 따라서 본 연구에서는 연령을 통제하고 추가적인 회귀분석을 실시하였다.

둘째로 고려해야 할 사항은 $\widetilde{\theta_i}$ 인데 이는 어떤 이유로 세대 내 출산을 결정하는 요인이 해당 기간 내에 변했을 경우이다. 예를 들어 $\widetilde{\theta_i}$가 높아진 세대가 체계적으로 일찍 혹은 늦게 출산환경이 좋은 곳으로 이주했을 경우이다. 본 연구는 방법론상 시간에 따른 환경이나 요인의 변화 등을 고려하기 어려운 구조인 것이 사실이다. 본 연구에서는 $\widetilde{\theta_i}$가 변화한 경우를 고려하기 위해 건강보험상 소득분위의 급격한 변동이나 가입형태(직장가입자, 지역가입자)의 변동이 생긴 관측치를 제외하고 회귀분석을 실행하였다.

이상의 논의를 바탕으로 추가적으로 실증분석한 결과를 <표 5>와 [그림 5]에서 제시하였다. 세 가지의 경우인데 앞서 설명한 대로 $\overline{\theta_i}$를 고려했을 때 연령 등으로 인해 $\delta_m \neq \delta_{m-1}$인 경우, 그리고 $\widetilde{\theta_i}$를 고려했을 때는 출산을 결정하는 요인이 해당 기간 내에 변했을 경우 등이다.

실증분석 방법으로는 전자의 경우 연령을 회귀식에 삽입하여 통제하였고(A. Age control) 후자의 경우 소득이 급격히 변한 관측치를 제거하거나 가입형태가 변한 관측치를 제거하였다. (D. Type change=0) 소득이 급격히 변한 관측치 제거는 다시 두 가지 경우, 즉 소득분위가 전체 20분위 중 한 번에 10분위 이상 급격히 변한 경우를 제거(B. Income change_case 1(10 grade change))한 경우와 5분위 이상 급격히 변한 경우를 제거(C. Income change_case 2(5 grade change))한 경우 등으로 나누어 실증분석하였다.[14)]

<표 5>와 [그림 5]를 통해 그 결과를 살펴보면 4가지 경우 모두 노출 연수 m과 b_m의 선형관계가 뚜렷하고 환경이 출산율에 미치는 영향이 본문의 실증분석 결과와 대동소이하여 앞의 결론을 뒷받침하였다.

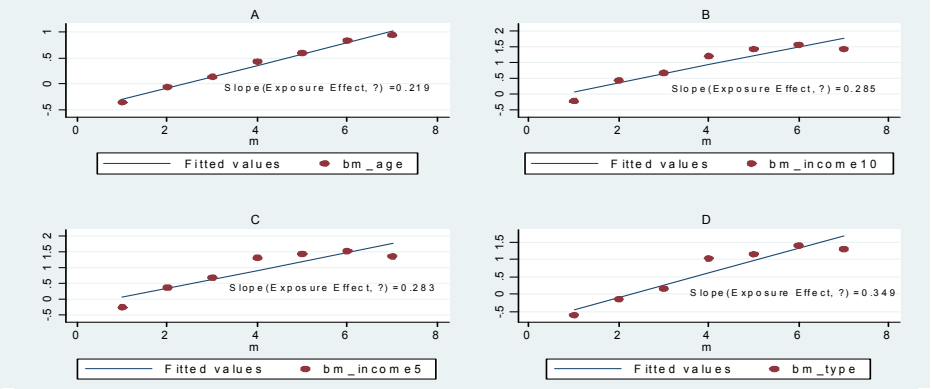
14) 사실 이 경우를 제외하고도 가입형태와 소득분위 등의 두 변수를 다양하게 변화시켜 실증분석을 실시하였지만, 그 결과는 대동소이하였다.

〈표 5〉 회귀분석결과

Variable	m=1	m=2	m=3	m=4	m=5	m=6	m=7
A. Age control							
b_m	-0.356 (0.038)	-0.063 (0.041)	0.132 (0.045)	0.426 (0.048)	0.591 (0.049)	0.831 (0.045)	0.940 (0.041)
age	-0.043 (0.000)	-0.054 (0.000)	-0.063 (0.000)	-0.067 (0.000)	-0.064 (0.000)	-0.059 (0.001)	-0.045 (0.000)
_cons	2.103 (0.051)	2.163 (0.056)	2.317 (0.061)	2.209 (0.066)	2.022 (0.067)	1.695 (0.063)	1.308 (0.058)
N	80118	87087	90617	93413	97741	116145	124641
B. Income change_case 1(10 grade change)							
b_m	-0.223 (0.057)	0.433 (0.061)	0.672 (0.066)	1.205 (0.070)	1.427 (0.070)	1.564 (0.064)	1.428 (0.057)
_cons	0.789 (0.073)	0.049 (0.078)	-0.160 (0.085)	-0.733 (0.090)	-0.922 (0.090)	-1.004 (0.082)	-0.674 (0.073)
N	44979	48217	49825	51506	54108	65401	71449
C. Income change_case 2(5 grade change)							
b_m	-0.248 (0.083)	0.375 (0.088)	0.695 (0.095)	1.322 (0.099)	1.445 (0.099)	1.534 (0.090)	1.367 (0.078)
_cons	0.786 (0.107)	0.057 (0.113)	-0.270 (0.122)	-0.974 (0.127)	-1.025 (0.128)	-1.033 (0.115)	-0.643 (0.101)
N	21261	22976	24305	25268	26731	33471	36965
D. Type change=0							
b_m	-0.598 (0.106)	-0.136 (0.108)	0.165 (0.116)	1.031 (0.119)	1.156 (0.123)	1.401 (0.107)	1.300 (0.094)
_cons	1.174 (0.135)	0.625 (0.138)	0.287 (0.149)	-0.732 (0.152)	-0.766 (0.157)	-0.956 (0.137)	-0.617 (0.121)
N	13062	14542	15431	15983	17080	22278	25145

주: ( ) 수치는 표준오차임, _cons는 상수항, N은 관측치.

〔그림 5〕 노출효과



주: Fitted values는 적합값, bm은 회귀계수, y축은 회귀계수의 크기, x축(m)은 노출연수, Slope(exposure effect, γ)는 노출효과를 나타내는 기울기.

다만 두 가지 차이점이 존재하는데 첫째, 그 기울기가 기본모형의 실증결과는 0.246이었는데 반해(A), 연령 통제한 경우가 이보다 낮은 0.219이고(B), 소득이 급격히 변화한 관측치 제거(10분위)의 경우 0.285(C), 소득이 급격히 변화한 관측치 제거(5분위)가 0.283이었으며(D), 가입형태가 변한 관측치 제거의 경우가 0.349로 가장 컸다. 환경의 영향이 더욱 뚜렷해지는 것이다. 둘째, (A)연령을 통제하였을 경우는 다른 실증결과에서 공통적으로 보이는 6년차와 7년차의 특성이 나타나지 않는다. 즉 다른 실증결과는 6년차에 이르러 환경의 영향이 사라지는 데 반해 연령을 통제한 경우는 b_m이 계속 증가하였다. 그 이유에 대해서는 추가적인 연구가 필요하다.

Ⅳ. 결 론

본 연구가 서론에서 했던 질문에 대한 완벽한 답이 되지는 못하더라도 연구의 출발점이 되길 바라는 마음에서 마지막으로 본 연구의 한계이자 향후 연구방향에 대해 두 가지를 언급하고자 한다. 첫째, 본 연구의 가장 큰 쟁점이자 추후 연구로서 보완되어야 할 부분은 가정 1의 검증문제와 이주시기에 따른 고정효과의 고려이다. 이는 추가적인 빅데이터 분석을 통해 가능할 것이다. 이는 추후 연구에서 더 고민해야 할 것이지만 예를 들어, 본 DB가 포함하고 있는 첫째아이와 둘째 아이 여부를 이용해 통제요소를 설정할 수 있을 것이다.

둘째, 출산과 환경에 관한 본 논문의 예비적 분석을, 정책적인 함의를 도출할 수 있는 본격적인 분석으로 발전시키기 위해서는 출산에 영향을 주는 환경적인 요소를 구체적으로 선정하고 장소효과와 선별효과를 구분하여 인과관계를 명확하게 밝히는 작업이 필요하다. Chetty and Hendren(2018a, 2018b)은 그들의 기회불평등 연구에서 세대 간 소득이동성의 지역적인 차이를 규명한 이후 탐색적인 연구를 통해 인종 등의 분리(Segregation)문제, 소득불평등, 공교육의 질, 편부, 편모 등의 가족구조, 사회적 자본 등을 그 지역적 차이의 원인으로 제시한 바 있다. 기존 연구에서 제시하고 있는 현금보조, 주택, 보육, 일자리 관련 정책이 출산율에 미치는 영향에 관한 연구를 본 논문의 방법론을 활용・발전시

켜 연구의 실용성과 정책적 함의를 제고할 수 있을 것이다. 2000년대 이후 각 지자체를 중심으로 활발히 도입된 출산장려금정책의 효과에 대한 연구를 본 연구의 방법론에 입각해서 수행할 수도 있다. 한 지역에 오래 거주한 여성들의 출산지수를 별도로 계산하여 현금지급정책이 그 지역의 출산율을 실제로 상승시켰는지 아니면 인근 지역으로부터의 이주를 유발하였는지를 구분하여 정책의 효과를 살펴보는 것이다.

사실 환경이 출산율에 중요한 영향을 미칠 것이라는 것은 어찌 보면 당연한 얘기일 지도 모른다. 본 연구의 의의는 이러한 상식을 과학적으로 그리고 엄밀하게 보여주려고 시도했다는 것과 이후의 분석 방향을 제시했다는 것이다. 특히 많은 관측치의 빅데이터를 이용하고 노출효과와 선별효과를 구분하여 실증함으로써 지속가능한 출산정책만이 시차를 두고 그 효과가 있을 수 있음을 주장하였다. 지자체의 무분별한 출산장려정책은 그것이 만약 장소효과나 노출효과보다 선별효과만을 주로 초래할 경우 일종의 인근 궁핍화에 불과해 정책의 효율성을 떨어뜨릴 것이다. 지역정책과 국가정책의 조율을 통해 좀 더 효율적인 정책을 설계할 필요가 있는 것이다.

참고문헌

강정구 · 마강래(2016).「지역의 주택가격이 초혼시기에 미치는 영향」.『한국지역개발학회지』29 (2) : 97~110.

김민영 · 황진영(2016).「주택가격과 출산의 시기와 수준 : 우리나라 16개 시도의 실증분석」.『보건사회연구』36 (1) : 118~142.

김우영 · 이정만(2018).「출산장려금의 출산율 제고 효과 : 충청지역을 대상으로」.『노동정책연구』18 : 61~98.

김준일(2021).「장기거주자의 출산율과 출산장려금 : 국민건강보험 빅데이터를 이용한 연구」.『Journal of the Korean Data Analysis Society』23 (3) : 1273~1284.

박창우 · 송헌재(2014).「출산장려금 정책이 출산에 미치는 영향 추정」.『응용

경제』 16 (1) : 5~34.

서문희 · 양미선 · 강기숙(2016).「보육과 출산의 연계성에 대한 거시-미시 접근」.『연구보고서 2016-44-05』 한국보건사회연구원, 한국보육진흥원.

송헌재 · 우석진(2015).「보육지원정책이 출산율과 여성 노동시장 참여율에 미친 거시적 성과 실증분석」.『재정정책논집』 17 (1) : 3~36.

이삼식 · 최효진(2011).「출산율 예측 모형 개발」.『한국인구학』 35 : 77~99.

이상협 · 이철희 · 홍석철(2016).「저출산대책의 효과성평가」.『연구보고서 2016-44-08』. 한국보건사회연구원, 경제추격연구소.

조영태 · 원성호 · 김수연 · 박준영(2018).「출산영향요인 발굴을 통한 미래 인구정책 방향」. 연구용역보고서. 국회사무처.

천현숙 · 김현표 · 정희남 · 김혜승 · 하수정 · 김진범 · 윤윤정 · 오민준 · 김태환(2012).『저출산 추세에 대응한 주택 및 도시정책 방향연구(Ⅰ)』. 국토연구원.

허만형(2020).「출산장려금의 출산율 제고 효과에 관한 연구 : 인구이동 매개효과 분석을 통한 시군구 비교분석」.『지방정부연구』 24 : 51~67.

Bergman, P., R. Chetty, S. DeLuca, N. Hendren, L. F. Katz and C. Palmer(2019). “Creating Moves to Opportunity : Experimental Evidence on Barriers to Neighborhood Choice”. NBER Working Paper, No. 26164.

Chetty, R. and N. Hendren(2018a). “The Impacts of Neighborhoods on Intergenerational Mobility I : Childhood Exposure Effects”. *The Quarterly Journal of Economics* 133 (3) : 1107~1162.

_____(2018b). “The Impacts of Neighborhoods on Intergenerational Mobility II : County-level Estimates”. *The Quarterly Journal of Economics* 133 (3) : 1163~1228.

Chetty, R., N. Hendren, P. Kline and E. Saez(2014). “Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States”. *The Quarterly Journal of Economics* 129 (4) : 1553~1623.

Chetty, R., N. Hendren and L. F. Katz(2016). “The Effects of Exposure to

Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment". *American Economic Review* 106 (4) : 855~902.

Drago, R., K. Sawyer, K. M. Shreffler, D. Warren and M. Wooden(2011). "Did Australia's Baby Bonus Increase Fertility Intentions and Births?". *Population Research and Policy Review* 30 (3) : 381~397.

Gauthier, A. H.(2007). "The Impact of Family Policies on Fertility in Industrialized Countries : A Review of the Literature". *Population Research and Policy Review* 26 (1) : 323~346.

Gauthier, A. H. and J. Hatzius(1997). "Family Benefits And Fertility : An Econometric Analysis." *Population Studies* 51 (3) : 295~306.

González, L.(2013). "The Effect Of a Universal Child Benefit On Conceptions, Abortions, and Early Maternal Labor Supply." *American Economic Journal : Economic Policy* 5 (3) : 160~188.

Katz, L. F., J. R. Kling, and J. B. Liebman(2001). "Moving to Opportunity in Boston : Early Results of a Randomized Mobility Experiment". *The Quarterly Journal of Economics* 116 (2) : 607~654.

Kling, J. R., J. B. Liebman, and L. F. Katz(2007). "Experimental Analysis of Neighborhood Effects". *Econometrica* 75 (1) : 83~119.

Milligan, K.(2005). "Subsidizing the Stork : New Evidence on Tax Incentives and Fertility". *Review of Economics and Statistics* 87 (3) : 539~555.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu(2013). "Long-term Neighborhood Effects on Low-income Families : Evidence from Moving to Opportunity". *American Economic Review Papers and Proceedings* 103 (3) : 226~231.

Oreopoulos, P.(2003). "The Long-run Consequences of Living in a Poor Neighborhood". *Quarterly Journal of Economics* 118 (4) : 1533~1175.

Abstract

A Research on Fertility Rate and Environment Using National Health Insurance Big Data : A Preliminary Analysis

Kim, Joonil

The purpose of this study is to analyze the relationship between fertility rate and environment using a new empirical approach. The existing approach on the effect of low fertility rate policy has limitations in empirical analysis without distinguishing the effect of place-specific environment and the selection effect from individual heterogeneity. Therefore, this study tries to devise a quasi-experimental research methodology using the National health insurance big data and contribute to establishing a more effective policy through measuring the environment by considering permanent residents and examining how the birth rate of movers is affected by the environment. As a result, this study shows that the environment of regions(Si-Gun-Gu) has a significant effect on the fertility rate, but that about 60% of the effects of good environment for childbirth in early stage are due to the effect of collecting women with a high chance of giving birth, and the effect of actually raising the birth rate gradually appeared after that, reaching the peak in about 6 years and then stagnating. This study has significance as a preliminary analysis for specific policy analysis.

Keywords : fertility rate, selection effect, place effect, exposure effect, quasi-experimental method, national health insurance big data