

초 록

본 논문에서 연구자는 다음의 세 가지 목표를 제안한다. 첫 번째, 저출산 문제의 명확한 진단이다. 2000년대 초중반의 저출산 담론과 2010년대의 저출산 담론은 서로 다른 문제 인식으로부터 출발해야 한다. 이를 위해 연구자가 분석에 사용할 데이터는 2006년과 2016년의 전국 단위 설문 자료인 한국종합사회조사(KGSS)이다. 특히 2016년 자료의 경우 이번 해 상반기에 공개됐기에 현재 혼인·출산·가족관에 대해서는 가장 최근의 정보를 담고 있다고 볼 수 있다. 두 번째, 단순히 기술적으로만 상황을 서술하려는 작업은 아니다. 본 논문은 **10년의 시간** 동안 어떻게 결혼·출산관이 변화했는지, 또 **성별**에 따라 어떻게 상이한 양상을 띠는지 그 차이에 구체적으로 주목한다. 이를 위해 교차효과(interaction effect) 분석 역시 실시할 것이다. 마지막으로 세 번째, 현실의 명확한 진단을 위해서 보다 개선된 계량적 방법론을 활용하고자 하였다. 이를 위해 연구자는 p-value 기반의 단순 최소자승추정법이 아닌 정규화 기법 LASSO를 적용해 변수 선택 방법에 기존의 사회과학 경험연구들과 차이를 두었다.

연구 대상은 한국의 미혼 남녀로, 일반적으로 연간 KGSS 조사에서 약 25% 정도를 차지한다. 본 연구에서는 지나친 저연령/고연령대 응답자를 제외시키기 위해 평균 $\pm 2 \times$ 표준편차로 연령 범위를 제한하였으며 그 결과 18세에서 44세 사이의 523명이 표본으로 선정되었다. 연구의 종속 변수는 ‘비혼/비출산/탈가족주의 가치관’이다. KGSS 2006/2016년도 데이터는 이와 관련해 총 여섯 개의 문항을 포함한다. 이들 전체가 혼인 여부 및 자녀 출산 여부를 직접적으로 묻지는 않지만 탈가족주의적 가치관이라는 잠재 변수(latent variable)로 통합될 수 있다고 판단했기에 문항들을 대상으로 회전된 주성분 분석(varimax rotated principal component analysis)을 실시, 단일 종속변수로 변경하여 활용하였다.

방법론적으로는 조정 LASSO(adaptive Least Absolute Shrinkage and Selection Operator) 회귀분석을 활용했다. 기존 사회과학 연구에서 주로 사용되어 온 OLS 회귀분석 대신 LASSO를 활용한 이유는 다음과 같다. 첫 번째, 예측의 분산값을 낮추는 정규화(regularization) 기법을 활용하기에 표본에의 과적합(overfitting)을 방지할 수 있다. 다시 말해 선택된 변수 및 모델이 연구자가 분석에 활용한 자료를 넘어서 일반적인 예측에 더 효과적이다. 두 번째, 최근 사회과학계에서 논란이 되고 있는 p-value 중심의 변수 선택방식에서 상대적으로 자유롭다. p-value는 ①자신의 연구 가설 그 자체를 검증하는 것이 아닌 빈도주의적 방식이기에 연구의 결과를 직관적으로 이해하기 어렵고 ②표본의 크기에 많은 영향을 받기에 표본 수가 클수록 ‘유의미한 연구결과’를 도출하기에 유리하고, 반대로 작다면 불리하다는 문제점이 있다. 세 번째, LASSO를 통해 도출한 회귀계수값은 고전적인 빈도주의(frequentist) 뿐만 아니라 베이즈주의의 관점에서도 해석 가능하다. OLS를 통해 도출한 회귀계수를 사전확률(prior)로, LASSO를 통해 도출한 회귀계수는 이중 지수분포를 경유한 사후확률(posterior)로 볼 수 있기 때문이다. 네 번째, 변수들 간의 다중공선성 문제를 해소할 수 있으며, 독립 변수의 정규성 가정으로부터 다소 자유롭기에 범용성이 크다.

분석 결과 및 결론은 다음과 같다. 첫 번째, 탈가부장적 가치관, 탈권위주의적 정치 성향이 비혼·저출산과 유의미한 관계를 맺는다는 직관적인 사실을 최신의 데이터를 바탕으로 재확인

할 수 있었다. 또한 이를 통해 구조적으로 저출산 가치관이 향후에도 계속 확산될 것이라는 다소 부정적인 예측 역시 가능했다.

두 번째, 교차효과를 활용한 조절 모형들은 시기와 성별에 따라 어떻게 혼인·출산관이 상이하게 나타나는지를 보여주었다. 특히 OLS와 달리 LASSO 정규화 기법을 활용함으로써 노동 시장에 참여한 여성들과 그렇지 않은 여성들 간의 차이를 유의미하게 확인할 수 있었다. 기존의 국내 연구자들 중 상당수는 한국에서 여전히 일과 가정이 양립 불가능하고, 또 노동시장에의 참여는 혼인과 출산을 저해하는 요인으로 기능한다고 주장해왔다. 그러나 2006년과 2016년 양자 간 데이터를 비교해본 결과, 2006년에는 해당 주장들에 부합하는 결과를 관찰할 수 있었지만 2016년에 들어서는 취업한 여성들이 취업한 남성과 마찬가지로 미취업자들에 비해 더 높은 결혼·출산 경향을 보이는 것으로 나타났다. 물론 이를 긍정적인 신호로만 해석할 수는 없다. 가령 미국과 일본에서는 노동 시장의 유연성 확대에 따른 중년 여성들의 재진입을 통해 긍정적이지 않은 방향으로 일과 가정의 양립이 실현되었기 때문이다. 이는 실질적인 일과 가족의 양립을 위해 제도적 지원을 확대한 결과로 해당 정책적 목적을 달성한 북유럽 사민주의 국가들과는 다른 양상이라고 할 수 있다. (신경아, 2010) 따라서 본 연구에서 확인된 ‘여성 노동시장의 임계점’을 지난 한국 사회가 어떻게 일과 가정의 양립을 지원해야 긍정적인 방식으로 혼인·출산의 확대를 추구할 수 있을지에 대해 향후 정책적인 논의가 필요할 것이다.

한국 사회의 비혼·저출산 가치관은 어떻게 성별과 연도에 따라 상이한가?

- 조정 LASSO 회귀분석을 통한 KGSS 2006, 2016년도 데이터 분석 -

강태영 (서강대학교 사회학과)

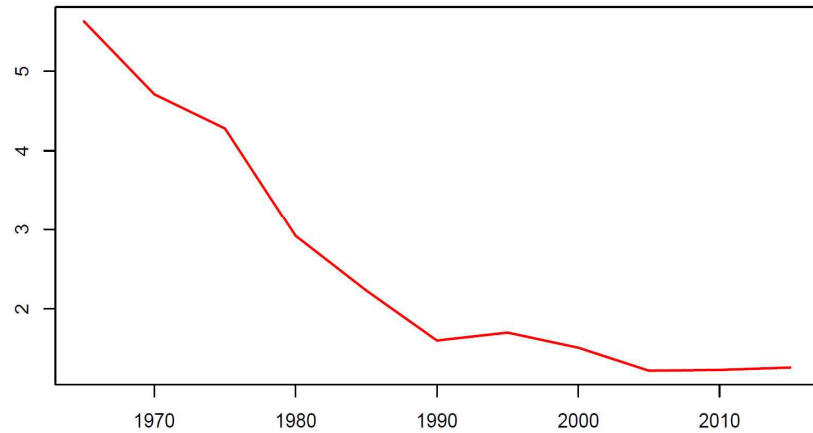
목차

1. 연구의 배경 및 목적
 2. 선행 연구 검토
 3. 연구의 설계 및 연구 방법
 4. 경험적 분석 결과
 5. 결론 및 이론적 함의
 6. 참고문헌
- 부록 : 연구재현을 위한 R 코드

1. 연구의 배경 및 목적

결혼과 출산, 취직과 연애 등 삶의 모든 것들을 포기했다는 자조적인 의미에서 활용된 단어인 ‘N포 세대’는 2010년대 청년들의 현실을 핵심적으로 관통한다. 그리고 사실 2000년대 중반부터 언급되어온 쟁점인 비혼과 저출산이라는 문제는 이러한 N포 청년들의 자연스러운 귀결이라고도 볼 수 있다. 다시 말해 담론으로서의 저출산은 오래되었지만 이는 근 몇 십년간 악화되어 온 청년들의 삶, 즉 경제적 생존을 위해 자신의 인적 자본에 투자하는 동시에 연애를 병행하고, 더 나아가 혼인과 출산에 필요한 비용까지 부담하는 역할을 동시에 수행할 수는 없는 사회 초년생들의 단면을 압축적으로 녹여내고 암울한 전망을 제시하기에 상당히 현재적인 개념이기도 하다.

한국의 평균 합계출산율이 인구 구조 유지를 위한 최소값인 2명에 미치지 못한지는 꽤나 오래되었다. 1980년대 후반에 이미 출산율이 평균 1.6명도 되지 않았기 때문이다. 이후 2000년대 초반 평균 1.1명이라는 유례 없이 낮은, 전세계적으로 최하위 수치를 기록하며 정부와 언론은 저출산을 주요한 사회 문제로서 호출하기 시작했다. 물론 10여년의 시간이 흐른 현재까지 상황은 전혀 개선되지 않았다. 그렇다면 과연 2000년대 초반 저출산 담론이 최초로 대두되고, 또 10년이 지난 현재 인구 절벽 문제와 함께 다시 등장할 동안 저출산 문제에 대한 우리의 인식과 대응은 어떻게 변해왔는가? 그리고 또 어떻게 변화해야할까?



[도표 1] 한국 합계출산율 기술통계표

이러한 문제에 답하기 위해 연구자는 다음의 세 가지 목표를 제안한다. 첫 번째, 문제의 명확한 진단이다. 2000년대 초중반의 저출산 담론과 2010년대의 저출산 담론은 서로 다른 문제 인식으로부터 출발해야 한다. 이를 위해 연구자가 분석에 사용할 데이터는 2006년과 2016년의 전국 단위 설문 자료인 한국종합사회조사이다. 특히 2016년 자료의 경우 이번 해 상반기에 공개됐기에 현재 혼인·출산·가족관에 대해서는 가장 최근의 정보를 담고 있다고 볼 수 있다. 두 번째, 단순히 기술적으로만 상황을 서술하려는 작업은 아니다. 본 논문은 **10년의 시간** 동안 어떻게 결혼·출산관이 변화했는지, 또 **성별**에 따라 어떻게 상이한 양상을 띠는지 그 차이에 구체적으로 주목한다. 이를 위해 교차효과(interaction effect) 분석 역시 실시할 것이다. 마지막으로 세 번째, 현실의 명확한 진단을 위해서 보다 개선된 계량적 방법론을 활용하고자 하였다. 이를 위해 연구자는 p-value 기반의 단순 최소자승추정법이 아닌 정규화 기법 LASSO를 적용해 변수 선택 방법에 기존의 사회과학 경험연구들과 차이를 두었다. 그럼 이제부터 한국 사회의 비혼·저출산 문제에 관해 어떠한 논의가 진행되어왔는지 알아보도록 하자.

2. 선행 연구 검토

혼인 및 출산에 관한 기존의 연구들은 크게 경제 성장과의 인과관계를 다루는 경제학적 접근과 20세기 후반 이후 주요 선진국들에서 대두된 저출산 현상에 대한 대응으로서의 인구학·사회학·사회복지학적 접근의 두 가지로 분류할 수 있다. 계량적 방법론 및 간학문적 연구들을 통해 개별 분과학문들이 통합되기에 양자를 이분법적으로 재단하기에는 어려운 것이 사실이다. 다만 편의상, 그리고 본 논문의 특성 상 연구자는 후자에 집중하였다. 우선 원론적인 수준에서 한국 내 비혼과 저출산의 원인으로 거론된 요인들은 정성호에 따르면 첫 번째, 전통적 가족관의 해체, 두 번째, 노동 시장의 악화, 세 번째, 여성의 노동 시장 참여에 따른 가사노동 및 출산 참여에의 유인 저하, 네 번째, 자녀 양육비/교육비 증대의 네 가지이다. (정성호, 2009)

첫 번째 요인으로 언급된 전통적 가족관은 단순히 “가족을 만들어야 한다는 의무의식의 상실” 정도로만 해석되기 쉽지만, 이는 혼인·출산 문제에서 복합적으로 작용해왔다. 가령 인구학자 Mindel Sheps에 의하면 한 가정에서 확률적으로 두 명의 아들을 얻기 위해서는 평균

적으로 3.88명의 자녀를 낳아야 한다. (Sheps, 1963) 그러나 90년대 이후로 한국 사회는 남아선호사상이 여전히 강하게 남아있음에도 불구하고 그보다 훨씬 낮은 초저출산 상태에 진입하게 되었다. Park과 Cho의 95년도 논문은 이러한 현상에 대해 중요한 의견을 제공한다. 그들에 의하면 아들을 낳기 위한 한국의 '선택적 낙태'는 보수적으로 계산했을 때에도 개별 가정들이 평균적으로 2.3~2.6명의 자녀를 갖게끔 만들었다. 그리고 이는 당연히 몇 십년 후의 결혼/출산 가능한 여성의 절대적 수의 감소에 상당한 영향을 끼쳤다. 다만 이들에 의하면 이러한 '신부 부족 현상'이 여성의 삶에 부정적인 것만은 아닌데, 상대적으로 연령이 높은 여성이 젊은 남성과 혼인할 기회가 증가하고, 또 그 희소성으로 인해 사회적 가치가 상대적으로 증가하기 때문이다. 그리고 이러한 여성의 가치 상승은 향후 세대에서 선택적 낙태 없는 평등한 출산의 신호로 작동할 수도 있다.(Park & Cho, 1995) 한편 이러한 남아선호사상은 기혼 여성의 출산관에게도 본질적인 영향을 끼친다. 가령 1991년도에 시행된 전국 가족 건강 조사에 기반한 다변량 분석에 의하면 첫째 아이로 아들을 낳은 여성은 굳이 둘째 자녀를 가지지 않으려 하는 경향이 관찰되었으며, 저자들에 의하면 이러한 유교적 영향력은 산모 개인의 종교관을 초월해서 나타난다. (Larsen et al, 1998) 이렇듯 전통적 가족관은 통념과 달리 해체 이전에도, 그리고 해체 이후에도 각기 다른 방식으로 절대적인 출생 자녀 수에 영향을 주는 것으로 나타난다.

두 번째와 세 번째 요인과 관련된 노동시장 상황에 대해서는 다음과 같은 설명들이 존재한다. Peter McDonald는 1960-1970년대를 거쳐 향상된 동아시아의 교육 수준이 여성의 사회 진출 기회를 확대시켰고, 1990년대 중후반의 동아시아 경제 위기로 인한 노동 시장 상황 악화가 청년들의 위험 회피 성향을 높였기에, 이들 두 요인이 저출산의 원인이 되었다고 주장한다. 그렇기에 그의 정책적 제언은 정부가 미혼 남녀들에게 기혼 부부에 대한 사회 보장 정책 확대 신호를 보내는 것이다. (McDonald, 2009) 이 때 여성의 노동 시장 참여 증가는 한국의 문화적 특수성과 결부되어 있는 중요한 쟁점으로 여겨진다. 자녀의 사회적 성공을 위한 노력과 결혼 후 가사노동에의 투신이 기혼 여성의 '자질'로 요구되는 한국의 유교적 특성상 여성들이 일과 가정 모두에 충실하는 것은 현실적으로 어렵기 때문이다. (Suzuki, 2008) 다층분석(multilevel analysis)을 통한 성평등주의와 여성의 출산에 관한 김영미의 분석은 그 연장선 상에서 매우 흥미로운 논의를 제공한다. 그에 의하면 일반적으로 여성의 취업과 성평등주의적 가치관은 출산과는 부정적인 관계를 맺는다. 그러나 전자의 경우 가족 대상 공적 서비스가 발전된 국가일수록, 후자의 경우 남녀 간 젠더 인식 차이가 적은 국가일수록 그러한 경향이 감소하며, 따라서 이는 한국이 여전히 일과 가정이 양립 불가능하다고 여겨지는 사회적 분위기에 대한 이론적 근거 역시 제공한다. (김영미, 2017) 또한 높은 사회 계층의 여성들은 결혼 이전에는 노동시장에 적극적으로 참여하나 결혼 이후에는 자녀 양육을 위해 노동 시장으로부터 이탈하는 것으로 나타난다. (김영미, 2014)

국가 간 데이터를 바탕으로 한 전반적인 추세 연구에서는 한국의 특수성과는 다른 양상이 관찰된다. 가령 22개 주요 선진국들을 대상으로 패널 분석을 실시한 결과 1980년대 이전까지 여성의 노동 시장 진출과 출산율은 음의 관계를 보였지만 1980년대를 거치며 관계가 역전되기 시작하고 1990년대부터는 오히려 여성이 적극적으로 노동에 참여할수록 출산율 역시 높게 나타나는 현상이 관찰된다. 여성의 어머니 역할과 노동자로서의 역할의 양립 불가능했던 과거의 상황이 개선되었기 때문이다. (Rindfuss et al, 2003) OECD 국가를 대상으로

한 분석 역시 1990년대 중반에 잠시 주춤하기는 했으나 1980년대 이후로 여성의 노동 시장 진출과 출산율은 계속적인 정(正)의 관계에 있는 것으로 나타난다. 그렇기에 논문의 저자인 Morgan은 선진국들의 입장에서 저출산이라는 문제는 자국의 자원으로 충분히 대응 가능하다는 낙관적인 의견을 제시한다. (Morgan, 2003) 김영미와 계봉오 또한 주요 국가들을 대상으로 분석 시 여성의 경제활동 참여율이 60% 이상인 국가에서는 양자가 정의 관계를 맺는다고 설명한다. 이를 바탕으로 저자들은 노동시장의 성 차별 개선과 양육 부담 감소의 두 가지 모두가 출산율 확대를 위한 필요조건이라고 주장한다. (김영미 & 계봉오, 2015)

마지막으로 네 번째, 자녀 양육의 경제적 부담 또한 젊은 부부들이 출산에 소극적이게끔 만드는 원인이다. 인터뷰를 바탕으로 한 질적 연구에 의하면 중산층 부모들은 여럿을 낳기 보다는 한 명에게 모든 자원을 투자해 입시 시장에서 자신의 자녀가 출세하기를 희망하는 것으로 나타났다. (김경근 & 최윤진, 2017) 그리고 이와 같은 자녀 교육 시장의 과열은 부모들로 하여금 자녀 출산에서 “양 보다는 질(quality over quantity)” 전략을 취하게 하는 유인이 된다. (Anderson & Kohler, 2013) 실제로 평균 양육비·교육비를 바탕으로 가구 당 자녀가 한 명 추가될 경우를 시뮬레이션 한 결과, 적자 가구의 비중이 약 1.7배에서 2배까지 증가하는 것으로 나타난다. (이상은 & 김희찬, 2017)

물론 저출산을 단기적으로 발생한 사회 문제라기 보다는 경제 성장 및 근대화에 의한 자연스러운 결과로 파악하는 연구 흐름도 있다. 장기 시계열 분석 하에서 비농업 가구와 화이트 칼라 직업군의 증가 및 교육 수준의 향상, 여성 문맹률의 감소는 출산율을 저하시키는 것으로 나타나기 때문이다. (Poston et al, 2009) 유사하게 대만과 한국, 일본 등 동아시아 저출산 국가들에서도 여성의 교육 수준 상승이 결혼 시장에서의 불리함을 낳았다고 분석된다. 이에 해당 저자는 증가하는 동거 문화의 확산과 비혼·동거 상태에서의 출산 확대의 두 가지를 혼인·저출산 문제의 현실적 대안으로 제시한다. (Jones & Gubhaju, 2009)

그렇다면 한국의 현재 상황은 어떠할까? 과연 저출산 대책은 적절하게 실행되고 있을까? 상황은 그리 긍정적이지 않다. Jones, Straughan과 Chan은 한국은 출산율이 지나치게 낮아져 인구 구조에 심각한 영향을 끼치는 시점이 되어서야 저출산 대책을 도입했기에 불안정하다고 지적했으며, 맞벌이가 증가하는 추세임에도 남성의 가사노동 참여율이 낮고 가족주의적인 저출산 대책의 수가 여전히 많은 점이 장애물로 기능한다고 주장했다.(Jones, et al, 2008) 엄동욱은 Lutz의 저출산함정(LFT, low fertility trap) 가설 (Lutz, et al, 2006)에 입각해 한국에서 ①출산율과 고령화 간에 음의 관계가 나타나고 ②실제 자녀 수가 이상 자녀 수를 하회하며 ③세대 간 상대소득 격차 역시 크게 나타나기에 저출산함정에 빠져 있으며 현실적으로 회복이 어렵다는 부정적인 진단을 내린다. (엄동욱, 2009) 연장선 상에서 서동희는 지난 10여년간의 저출산 대책의 실효성이 사실상 없었음을 인정하고 이제는 현실적으로 외국인 노동자 유입과 여성·고령층 노동력의 확보 등의 ‘저출산 적응정책’으로 정책적 방향 전환을 해야 한다고 주장한다. (서동희, 2015) 반면 일본의 경우 여성의 취업시장 진출 확대를 혼인·출산과 대립되는 개념으로 파악하지 않고 오히려 가정과 일의 양립이라는 구조적 문제에 초점을 두어서 정책을 진행하는, 한국과는 다른 모습을 보여주었다. 그렇기에 이는 형식적인 육아휴직제도 및 보육지원액 확대만을 추구하는 한국의 저출산 정책과는 대비된다고 평가받기도 한다. (Kang & Song, 2017)

3. 연구의 설계 및 연구 방법

본 연구에 활용할 데이터는 가족관, 성 역할 가치관, 가부장제에 관해 풍부한 문항을 포함하고 있는 2006년부터 2016년까지의 누적 한국종합사회조사(KGSS, Korean General Social Survey)이다. 한국종합사회조사는 2003년부터 시행되어 오고 있는 연간 사회조사로, 한국노동패널과 함께 국내에서 가장 양질의 결과를 도출해온 설문 데이터로 평가 받고 있다. 현재 KGSS는 성균관대학교 서베이 리처치 센터를 중심으로 시행되고 있으며 랜덤 샘플링 및 반복 조사의 원칙에 충실하다는 장점을 지닌다.

연구 대상은 한국의 미혼 남녀로, 일반적으로 연간 KGSS 조사에서 약 25% 정도를 차지한다. 본 연구에서는 지나친 저연령/고연령대 응답자를 제외시키기 위해 평균±2×표준편차로 연령 범위를 제한하였으며 그 결과 18세에서 44세 사이의 523명이 표본으로 선정되었다. 연구의 종속 변수는 ‘비혼/비출산/탈가족주의 가치관’이다. KGSS 2006/2016년도 데이터는 이와 관련해 총 여섯 개의 문항을 포함한다. 이들 전체가 혼인 여부 및 자녀 출산 여부를 직접적으로 묻지는 않지만 탈가족주의적 가치관이라는 잠재 변수(latent variable)로 통합될 수 있다고 판단했기에 문항들을 대상으로 회전된 주성분 분석(varimax rotated principal component analysis)을 실시, 단일 변수로 변경하여 활용하였다. 6개 종속변수들의 내용은 다음과 같다.

종속변수명	사용 설문 문항
비혼/비출산/탈가족주의 가치관	남편은 부인보다 나이가 많지 않아도 된다
	결혼하더라도 아이를 가질 필요가 없다
	결혼하지 않은 사람은 결혼한 사람보다 행복하다
	결혼할 의사가 없이 함께 사는 것도 괜찮다
	이혼하고 싶다면 자녀가 장성할 때까지 기다릴 필요는 없다
	결혼 생활이 원만하지 못하면 이혼이 최선책이다

[도표 2] 종속변수 활용 문항

통제 변수는 ①성별, ②지역, ③연령, ④종교, ⑤가족교육수준, ⑥취직 여부, ⑦월간가계소득, ⑧설문시행연도이다. 성별과 지역, 종교 신자 여부와 취직 여부는 더미 변수로 처리하였다. 가족교육수준의 경우, 학제 구분을 바탕으로 한 학력 응답을 교육년수로 환산해 연속변수로 변경하였으며 월간가계소득의 경우 초고소득자로 인한 분석 편향을 예방하고자 평균±2×표준편차로 범위를 제한, 월소득 최대 1139만원인 응답자들만 분석 대상으로 포함시켰다. 독립 변수로는 ①계급상향인식, ②가부장적 가치관, ③우파권위주의성향(RWA), ④정치적 효용감, ⑤경제적 만족도의 총 다섯 가지를 활용했으며 이 중 계급상향인식을 제외한 모든 독립 변수들은 리커트 척도로 측정되었다. 또한 변수 간 교차효과(interaction effect)를 파악하기 위해 연도와 성별을 중심으로 주요 독립변수들과의 교차항을 추가적으로 사용하였다. 주요 통제/독립변수들의 설문 문항 내용은 다음과 같다.

독립변수명	사용 설문 문항
계급 상향 인식	한국사회의 최하층을 1로 하고 최상층을 10으로 한다면 귀하는 어디에 속한다고 생각하십니까?
	귀택의 월평균 총소득은 대략 어느 정도 됩니까?
가부장적 가치관	아내는 남편을 내조하는 것이 더 중요하다
	아내의 역할은 가정과 가족을 돌보는 것이다
	불경기에는 여자를 우선적으로 해고해도 된다
	아버지의 권위는 어떤 경우에도 존중되어야 한다
	자식은 부모에게 명예가 되는 일을 해야 한다
	가계 계승을 위한 아들이 있어야 된다
우파 권위주의 성향(RWA)	극단적인 집회는 허용되어서는 안 된다
	극단적인 서적 출판은 허용되어서는 안 된다
	정부에 항의하는 집회는 허용되어서는 안 된다
	항의 시위와 데모는 허용되어서는 안 된다
가족교육수준	귀하의 최종학력은 어떻게 됩니까?
	귀하의 어머니의 최종학력은 어떻게 됩니까?
	귀하의 아버님의 최종학력은 어떻게 됩니까?

[도표 3] 독립변수 활용 문항

이 중 ‘계급 상향 인식’은 응답자들의 월간 가계소득을 10분위로 분할한 후, 주관적으로 체감하는 10분위수와의 차이, 즉 “특정 행위자가 얼마나 스스로의 경제적·계급적 위치를 실제보다 상향 혹은 하향평가 하는지”를 계산하여 만든 변수이다. 따라서 그 값은 ‘주관 10분위 - 객관 10분위’의 정수 꼴로 나타난다. ‘우파 권위주의 성향(Right Wing Authoritarianism, RWA)’은 심리학자 알트마이어에 의해 제안된 개념으로, 그에 의하면 사회적 학습의 결과물과 성격, 삶에서의 경험 등을 복합적으로 측정할 수 있다. (Altemeyer, 1981) 예시 문항은 다음과 같다.

우파 권위주의 성향(RWA) 측정 예시 문항
<ol style="list-style-type: none"> 1. 현재 설립된 권위는 일반적으로 올바르다고 볼 수 있으며, 반면에 급진주의자들과 저항론자들은 그저 “목소리만 크게(loud mouths)” 널 뿐이다. 2. 우리나라가 위기를 극복할 수 있는 방법은 전통적인 가치로 돌아가고, 강한 지도자를 내세워 불평분자들을 제거하는 것이다. 3. 삶에서 한 가지 방식만 존재하는 것은 아니다. 모두가 각자의 길을 걸어가야 한다.

[도표 4] 우파권위주의성향 측정 예시 문항

우파 권위주의 성향(RWA)은 정치적, 사회적 가치관과 밀접한 관계를 맺고 있다. 가령 아동에 대한 육체적 훈육을 선호하는 부모일수록 보수적인 대통령을 선호했으며 (Hetherington,

2009) 전반적으로 높은 RWA 수치를 갖는 행위자들은 사회적 순응성이 높은 동시에 생활 세계를 위험한 공간으로서 인식하고 있었다. (Duckitt, 2006) KGSS 설문지가 RWA 관련 문항들을 직접적으로 측정하지는 않지만 시위 및 집회에 관한 의견들을 바탕으로 해당 변수를 변형·재구성하였다.

연구의 모델은 총 세 가지이며 각 모델마다 위계적으로 변수들을 추가시켰다. 첫 번째 모델에는 통제변수 및 주요 독립변수들을 그대로 투입했으며 두 번째 모델에는 성별(남성)과 취업여부/가부장적가치관/종교신자여부 간의 교차항들을, 세 번째 모델에는 연도(2016년)과의 교차항들을 더하였다. 또한 효과 크기(effect size)를 보다 수월하게 이해할 수 있도록 모든 독립변수들은 이항변수의 경우 더미변수로 변환하고 연속변수들은 Z-표준화시켰으며, 종속변수 값은 0에서 100사이의 값으로 재척도화했다.

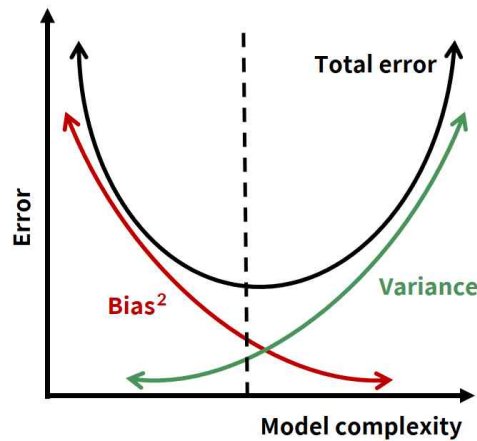
방법론적으로는 조정 LASSO(adaptive Least Absolute Shrinkage and Selection Operator) 회귀분석을 활용했다. 기존 사회과학 연구에서 주로 사용되어 온 OLS 회귀분석 대신 LASSO를 활용한 이유는 다음과 같다. 첫 번째, 예측의 분산값을 낮추는 정규화(regularization) 기법을 활용하기에 표본에의 과적합(overfitting)을 방지할 수 있다. 다시 말해 선택된 변수 및 모델이 연구자가 분석에 활용한 자료를 넘어서 일반적인 예측에 더 효과적이다. 두 번째, 최근 사회과학계에서 논란이 되고 있는 p-value 중심의 변수 선택방식에서 상대적으로 자유롭다. 우선 원론적인 이야기부터 해보자면, p-value는 ①자신의 연구 가설 그 자체를 검증하는 것이 아닌 빈도주의적 방식이기에 연구의 결과를 직관적으로 이해하기 어렵고 ②표본의 크기에 많은 영향을 받기에 표본 수가 클수록 ‘유의미한 연구결과’를 도출하기에 유리하고, 반대로 작다면 불리하다는 문제점이 있다. 물론 근 몇 년간 p-value/영가설검정을 둘러싼 논쟁들은 연구의 재현성에 초점을 둔다. 가령 버지니아 대학의 브라이언 노섹이 100개의 영가설검정 기반 심리학 실험을 재현한 결과 이들 중 재현 가능한 경우는 약 3분의 1 밖에 되지 않았다. (Open Science Collaboration, 2015) 이에 사회심리학 저널인 <Basic and Applied Social Psychology>는 p-value를 포함한 논문은 더 이상 출간하지 않을 것이라고 선언하기도 했다. (Trafimow & Marks, 2015) 또한 지난 2016년 3월에는 전미통계학회(ASA) 역시 “분석 결과의 유효성을 p-value 0.05라는 관례적 수치만으로 판단하는 것은 옳지 않다”는 내용의 성명을 발표했다. (Wasserstein & Lazar, 2016) 이렇듯 p-value에 기초한 기계적인 유의미성 판단 방식은 최근 많은 논란을 야기했으며 그 해결책으로 ①신뢰구간의 병기, ②계수 추정치의 크기를 중심으로 실질적 유의미성 표기, ③베이저안(Bayesian) 통계 기법의 사용 등이 논의되고 있다. 물론 연구자가 활용한 LASSO 역시 이와 함께 거론되는 대안이다. (박종희, 2015) 세 번째, LASSO를 통해 도출한 회귀계수값은 고전적인 빈도주의(frequentist) 뿐만 아니라 베이즈주의의 관점에서도 해석 가능하다. OLS를 통해 도출한 회귀계수를 사전확률(prior)로, LASSO를 통해 도출한 회귀계수는 이중 지수분포를 경유한 사후확률(posterior)로 볼 수 있기 때문이다. 네 번째, 변수들 간의 다중공선성 문제를 해소할 수 있으며, 독립 변수의 정규성 가정으로부터 다소 자유롭기에 범용성이 크다. 그렇다면 이제부터 LASSO가 어떻게 변수를 선택하는지, 그 과정에 대해 간략하게 알아보자.

최소자승추정법(OLS, Ordinary Least Square)의 경우 잔차제곱합(RSS)을 최소화하는 방식으로 회귀 계수를 추정한다. 반면 LASSO회귀 분석은 이 RSS에 L_1 페널티(penalty, 모델의 정확성을 위해서는 이 값을 줄여야 한다)를 추가해서 회귀 모델을 정규화(regularize)한다. 이를 수학적으로 표현하면 다음과 같다. (Tibshirani, 1996)

$$OLS : \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2$$

$$LASSO : \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j|$$

L_1 penalty의 λ 는 조율 모수(tuning parameter)라고 부르며 변수들 간의 상호작용을 제한하고 그 값이 커질수록 회귀계수추정치 of 분산을 감소시킨다.¹⁾ 물론 L_1 penalty 자체의 값은 작아져야 하기에 λ 가 무한하게 증가하지는 않으며 분산과 편향(bias)이 절충되는 지점에서 그 값이 결정된다. 바로 [도표 3]의 편향-분산 트레이드오프 그래프에서 점선으로 표현된 ‘모형 복잡도가 최적화되는 지점(optimum model complexity point)’에서 말이다.²⁾



[도표 5] 편향과 분산의 상충관계

L_1 penalty가 작아지기 위해서는 β_1 부터 β_p 까지의 계수들을 0에 가깝도록 만들어야 하며 이 수축 (shrinkage) 과정은 다음의 제한 영역(constraint area, 아래 식에서는 t 로 표현됨)에 관한 식으로 표현할 수 있다.

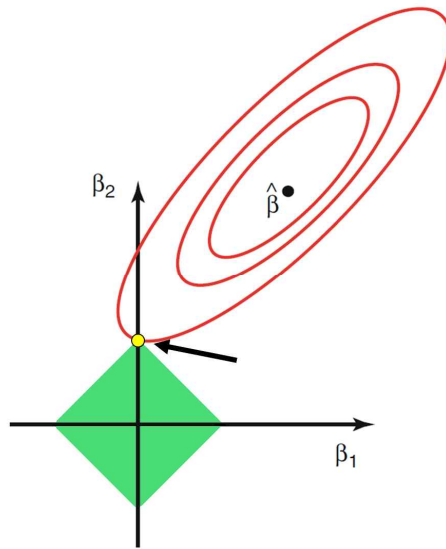
$$\operatorname{argmin}_{\beta_1, \dots, \beta_p} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \text{ subject to } \sum_{j=1}^p |\beta_j| \leq t$$

제한 영역의 개념을 수식으로만 이해하기 어렵다면 아래의 그림을 통해 보다 직관적으로 이해해보자. 다음은 계수가 두 개만 있는 상황(β_1 과 β_2)을 가정했을 때, LASSO의 계수 축소

1) 일반적인 OLS 회귀 모델은 편향(bias)은 작지만 분산(variance)이 큰 편이며, 데이터의 노이즈까지 패턴으로 해석해버리는 경향이 있다. ([도표 5]의 오른쪽) 이 때 LASSO 정규화를 실시할 경우 편향은 미세하게 증가하지만 분산은 획기적으로 감소하게 되어 모델의 정확성이 높아진다. ([도표 5]의 중앙)

2) <http://scott.fortmann-roe.com/docs/BiasVariance.html>의 그래프를 편집 및 수정하였다.

방식을 시각화한 것이다. 제한 영역은 두 개의 β 계수에 절댓값을 씌운 형태, 즉 $|\beta_1| + |\beta_2|$ 이기에 그림에서 볼 수 있듯 초록색 마름모꼴로 나타난다. OLS 추정치인 $\hat{\beta}$ 을 둘러싼 타원의 등고선은 잔차제곱합(RSS) 영역을 좌표평면에 표현한 것이다.³⁾



[도표 6] 2차원 LASSO 시각화

[도표 4]에서 화살표로 표시한 노란 점은 등고선이 제한 영역의 첨점에서 만나는 곳으로, 이 지점에서 설명력이 덜한 변수의 계수 추정치(β_1)는 완전히 0이 되고 β_2 만 최종 모델에 포함된다. (\therefore 노란 점의 β_1 축 좌표는 0이다.) 정리하자면, LASSO의 계수 추정치를 축소하는 과정에서 계수의 크기가 아주 작은 경우 그 값이 아예 0으로 수렴하며 최종 모델에서는 이들을 제외한 변수들이 선택(variable selection)된다.

다만 LASSO의 변수 선택 방식이 일관적이지 않은 경우도 종종 존재하며, 이에 조정 가중 벡터(adaptive weights vector)를 포함하는 조정 LASSO(adaptive LASSO)가 그 대안으로 제시되었다. (Zou, 2006) 가중 조정 벡터인 $\hat{\omega}_j$ 를 기존 LASSO에 추가한 식은 다음과 같다.

$$\text{adaptive LASSO} : \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \hat{\omega}_j |\beta_j|$$

가중 조정 벡터란 일종의 추가적 페널티로, 최초 회귀 계수 추정치(initial estimate of coefficient)인 $\hat{\beta}_j^{ini}$ 에 양의 상수 γ ⁴⁾을 제공하여 역수를 취한 값이다. 수식으로는 아래와 같이 표현할 수 있다.

$$\hat{\omega}_j = \frac{1}{(|\hat{\beta}_j^{ini}|)^\gamma}$$

3) RSS는 회전된 형태의 타원 방정식과 같다

4) 논문의 저자인 Zou는 0.5나 1, 2를 사용할 것을 제안한다. 연구자의 경우 1을 사용했다.

관례적으로 최초 회귀 계수 추정치 $\hat{\beta}_j^{ini}$ 로는 능형회귀(ridge regression)⁵⁾를 통해 구한 값을 활용하며 연구자 역시 이와 같은 방법을 사용했다.

정리하자면, LASSO 회귀분석은 ①회귀 계수 축소를 통해서 영가설검정과는 다른 방식으로 변수 선택이 가능하고 ②그 과정에서 영향력이 적은 변수들은 자동으로 계수값이 0으로 축소되며 ③분산을 줄이기에 과적합을 방지해 모델의 예측성을 향상시킬 수 있다는 특징을 지닌다. 뿐만 아니라 본 연구가 활용하는 데이터의 수는 500여명이기에 OLS를 사용할 경우 p-value가 과대 측정되는 문제가 발생하기 쉽다. 위와 같은 이유들로 인해 본 연구는 조정 LASSO 회귀분석을 시행하게 되었다. 물론 변수 선택 방식의 비교를 위해 OLS 회귀분석 역시 함께 제시하였다.

4. 경험적 분석 결과

표본 응답자들의 기초적인 인구통계적 정보는 다음과 같이 나타났다. 성별과 종교 신자 여부, 취직여부는 절반에 가깝게 나뉘어있었으며 44세 미만의 미혼 남녀 집단의 특성 상, 청년 거주자가 적은 강원도와 제주도 출신 응답자는 적게 포함되어 있었다.

변수	응답범주	빈도(%)	변수	응답범주	빈도(%)
성별	남성	57.17%	종교신자여부	종교있음	46.85%
	여성	42.83%		종교없음	53.15%
거주지역	서울	23.33%	취직여부	취직	49.52%
	경기도	19.89%		미취직	50.48%
	강원도	3.82%	응답년도	2006년	63.29%
	충청도	9.56%		2016년	36.71%
경상도	35.18%				
전라도	7.07%				
	제주도	1.15%			

[도표 7] 인구학적 통제 변수 기술통계

‘계급 상향 인식’ 변수의 근간이 되는 월간 가계소득 10분위 (좌측) 및 주관적 체감 소득 10분위 (우측)의 기술 통계적 정보는 아래의 표와 같다. 가계소득 10분위의 경우 동일한 구간 길이를 설정하고자 하였으나 응답 금액이 불균등하게 분포되어 있기에 일부 중위 구간(5, 6, 8분위)에서는 불가피하게 범위 차이가 존재한다. 분석 결과, 변수 가공 이전 단계의 ‘주관적 체감 소득 10분위’와 연구자가 가공한 ‘계급상향인식’ 변수는 모두 정규분포에 가까운 모습을 보였다.

5) 능형 회귀란 LASSO와 매우 유사한 개념으로, LASSO의 L_1 penalty 대신 L_2 penalty를 사용하며 제한 영역은 절댓값이 아닌 제곱 형태, 즉 2차원 좌표평면 상에서는 마름모꼴이 아닌 원의 형태로 나타난다. 그 이외의 자세한 사항은 다음의 논문을 참고. (Hoerl and Kennard, 1970)

구간	객관소득분위	빈도(%)	주관적 체감 소득 분위	빈도(%)
~65만원	1	10.10%	1	3.40%
66~140만원	2	9.97%	2	4.48%
141~199만원	3	9.85%	3	11.60%
200~225만원	4	10.42%	4	21.06%
226~275만원	5	9.59%	5	26.92%
276~300만원	6	11.39%	6	17.78%
301~390만원	7	7.27%	7	11.54%
391~475만원	8	11.45%	8	2.52%
476~585만원	9	8.69%	9	0.50%
586만원~	10	11.26%	10	0.19%

[도표 8] 2006년도 월간 가계소득 및 주관적 체감 소득 10분위

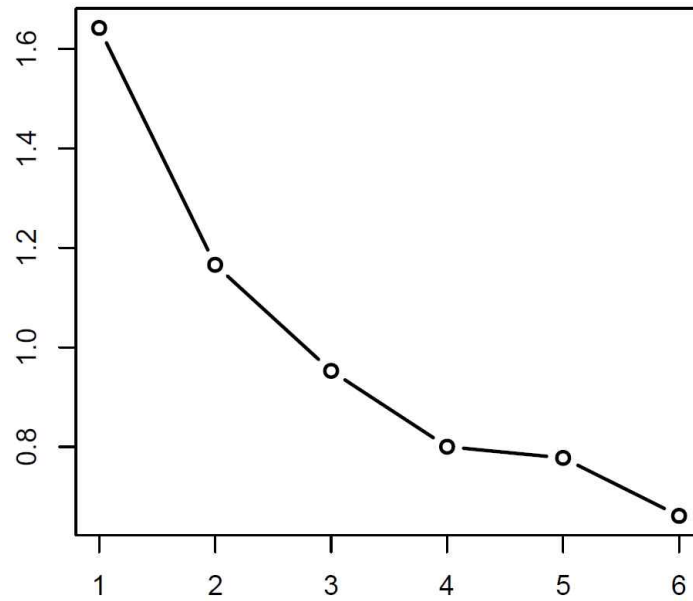
구간	객관소득분위	빈도(%)	주관적 체감 소득 분위	빈도(%)
~50만원	1	10.48%	1	4.21%
51~128만원	2	9.13%	2	4.21%
129~200만원	3	12.66%	3	12.33%
201~295만원	4	7.47%	4	16.06%
296~325만원	5	7.88%	5	27.53%
326~400만원	6	11.83%	6	19.12%
401~500만원	7	13.28%	7	12.43%
501~695만원	8	8.40%	8	3.06%
696~983만원	9	9.44%	9	0.29%
984만원~	10	9.44%	10	0.76%

[도표 9] 2016년도 월간 가계소득 및 주관적 체감 소득 10분위

계급상향/하향인식 정도	빈도(%)
[-9, -8]	0%
[-7, -4]	16.83%
[-3, -1]	38.24%
0	13.58%
[+1, +3]	23.9%
[+4, +7]	7.26%
[+8, +9]	0.19%

[도표 10] '계급상향인식' 변수 기술통계표

주요 독립변수 및 종속변수에 대한 주성분 분석(PCA, Principal Component Analysis)을 실시한 결과, 다음과 같은 값을 얻을 수 있었다. 주성분 추출에는 베리맥스 회전(varimax rotation)을 활용했으며 이후 표준화 적재값을 각 변수에 곱하여 단일 요인으로 변환 후 분석에 사용했다.

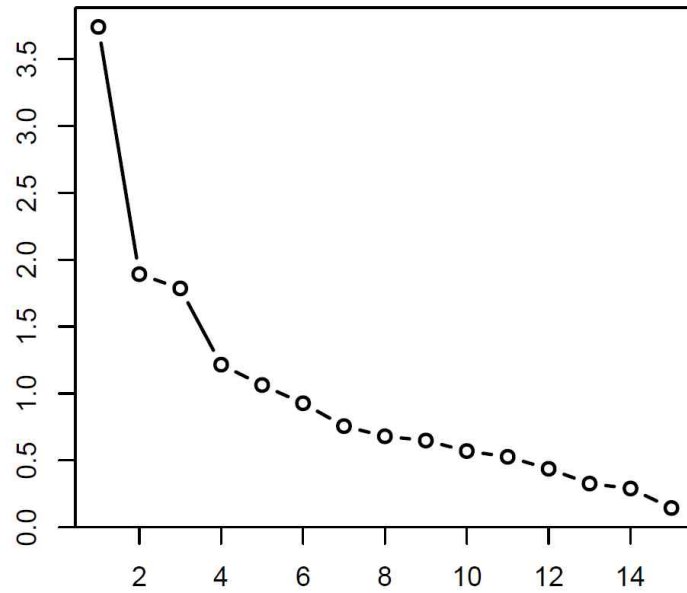


[도표 11] 종속변수 주성분분석 스크리 도표

변수명	표준화된 적재값
남편은 부인보다 나이가 많지 않아도 된다	0.470
결혼하더라도 아이를 가질 필요가 없다	0.614
결혼하지 않은 사람은 결혼한 사람보다 행복하다	0.553
결혼할 의사가 없이 함께 사는 것도 괜찮다	0.469
이혼하고 싶다면 자녀가 장성할 때까지 기다릴 필요는 없다	0.540
결혼 생활이 원만하지 못하면 이혼이 최선책이다	0.476

[도표 12] 종속변수 주성분분석 결과

종속변수인 ‘비혼/저출산/탈가족주의 가치관’의 경우, 주성분분석에 따른 잔차제곱근합 (RMSR)은 0.1596으로, 한 개 주성분의 활용의 χ^2 값에 대한 가설 검정 결과는 유의확률 (p-value)은 4.694×10^{-88} 으로 나타났다.

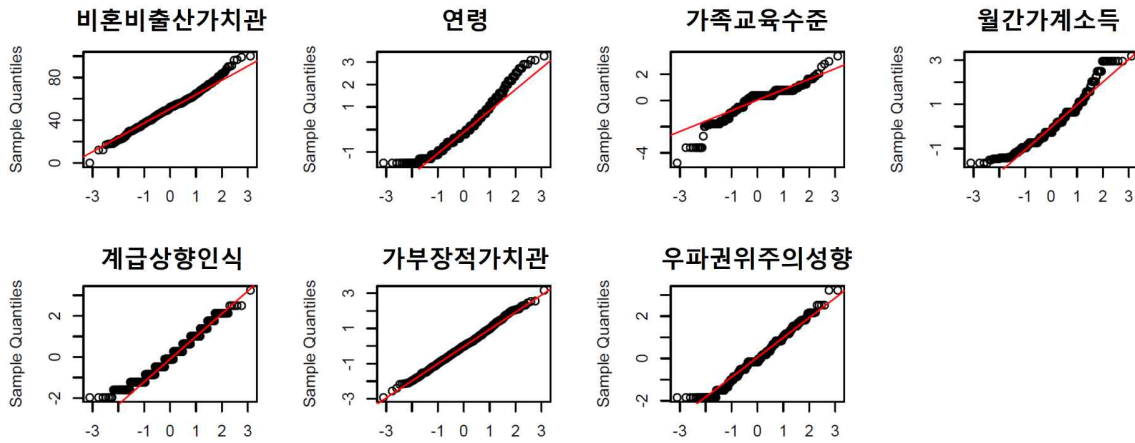


[도표 13] 연속형 독립변수 주성분분석 스크리 도표

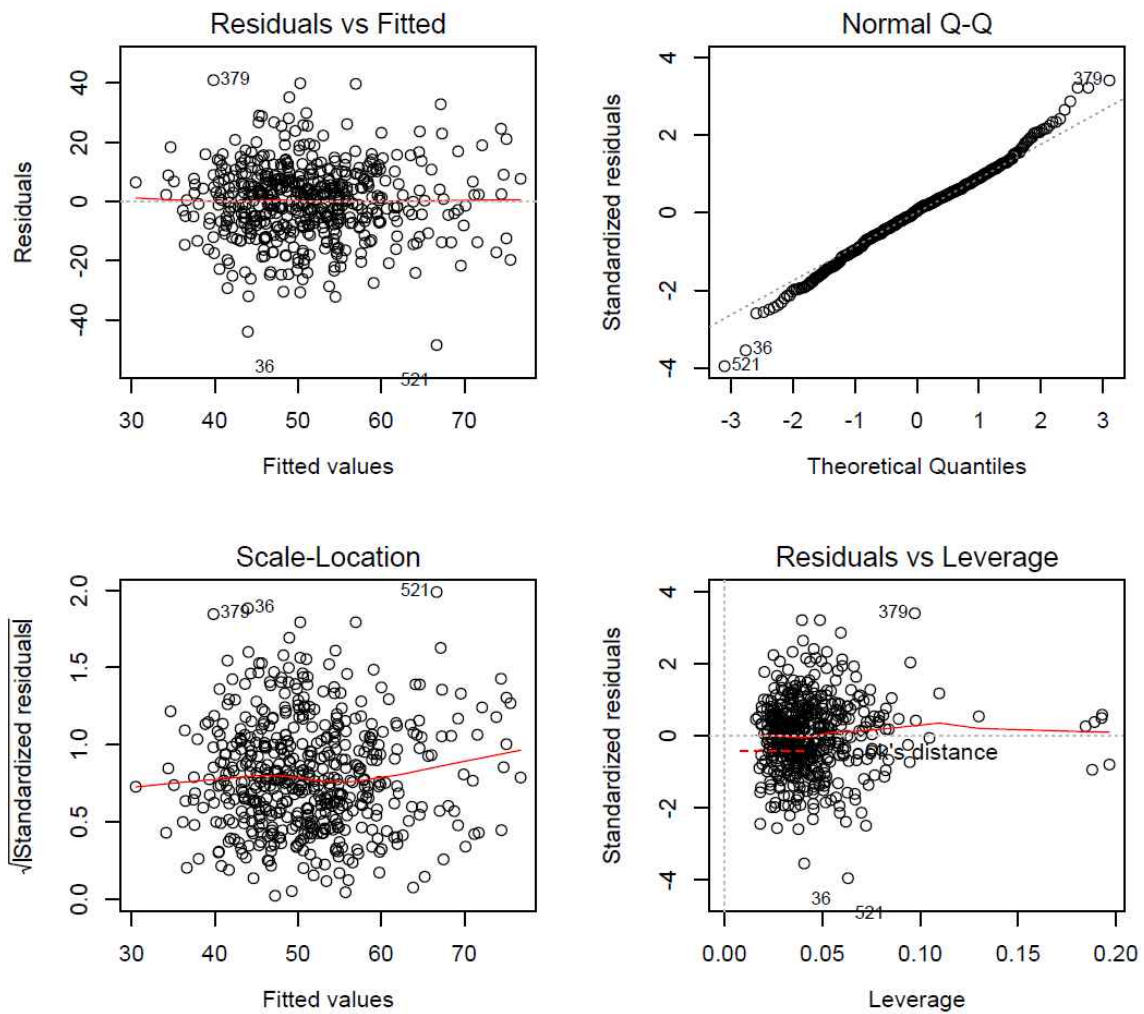
요인	변수명	표준화된 적재값		
가부장적 가치관	아내는 남편을 내조하는 것이 더 중요하다	0.762		-0.197
	아내의 역할은 가정과 가족을 돌보는 것이다	0.709	0.141	-0.205
	불경기에는 여자를 우선적으로 해고해도 된다	0.617		
	아버지의 권위는 어떤 경우에도 존중되어야 한다	0.580	0.119	
	자식은 부모에게 명예가 되는 일을 해야 한다	0.602		
	가계 계승을 위한 아들이 있어야 된다	0.731	0.106	
	자신보다 가족의 안녕과 이해를 우선시해야 한다	0.538	0.186	
우파 권위주의 성향	극단적인 집회는 허용되어서는 안 된다		0.836	
	극단적인 서적 출판은 허용되어서는 안 된다		0.821	
	정부에 항의하는 집회는 허용되어서는 안 된다	0.212	0.612	
	항의 시위와 데모는 허용되어서는 안 된다	0.161	0.571	-0.104
가족 교육수준	귀하의 최종학력은 어떻게 됩니까?	-0.105		0.919
	귀하의 어머니의 최종학력은 어떻게 됩니까?			0.885
	귀하의 아버님의 최종학력은 어떻게 됩니까?			0.617
-	계급상향인식	0.176		

[도표 14] 주요 독립변수 주성분분석 결과

인구 통제 변수들을 제외한 독립 변수들을 대상으로 실시한 주성분분석에서는 잔차제곱근합 (RMSR)이 0.0829로, 세 개 주성분의 활용의 χ^2 값에 대한 가설 검정 결과는 유의확률 (p-value) 1.17×10^{-13} 으로 나타났다. 계급상향인식은 어떠한 주성분에도 포함되지 못해 단일 변수 그대로 사용하였다.



[도표 15] 정규성 검정을 위한 연속형 변수들의 분위수-분위수 도표(Q-Q plot)



[도표 16] 회귀분석모형 잔차 진단

[도표 15]와 [도표 16]⁶⁾은 연속형 독립변수들 및 회귀모형 잔차의 정규성 진단 결과이다. 연속형 변수들은 모두 정규분포에 가까웠으며 회귀모형의 잔차 역시 이상치(outlier)로부터 특별히 영향 받지 않았으며 정규성을 충족하는 것으로 관찰된다. 이분산성(heteroskedas-

6) 도표 18의 Model 3 기준

ticity) 진단을 위한 Studentized Breusch-Pagan 검정을 실시한 결과, 95% 유의 수준에서 영가설은 기각되지 못했으며 이를 통해 우리는 회귀모델의 잔차들이 동분산(homoskedasticity) 가정을 유지한다고 추정할 수 있다. 종합하자면 본 연구의 회귀모델은 변수들의 분포 및 모델 잔차들이 정규성을 충족하며, 동분산을 갖기에 다중회귀분석을 실시하기에 적합하다고 볼 수 있다.

Regression Model	χ^2	degree of freedom	p-value
Model #1	12.133	16	0.7348
Model #2	13.248	19	0.8256
Model #3	23.581	22	0.3696

[도표 17] 회귀모델에 대한 Studentized Breusch-Pagan 이분산 검정 결과

(다음 페이지에서 계속)

OLS regression	Model #1	Model #2	Model #3
Intercept	56.411*** (0.000)	55.504*** (0.000)	52.921*** (0.000)
연령	2.212** (0.002)	2.335** (0.001)	2.321** (0.001)
경기도 거주	-0.499 (0.777)	-0.784 (0.656)	-1.109 (0.528)
강원도 거주	-4.397 (0.163)	-3.991 (0.204)	-3.645 (0.243)
충청도 거주	-2.487 (0.266)	-2.34 (0.298)	-1.937 (0.392)
경상도 거주	0.458 (0.763)	0.313 (0.836)	0.662 (0.667)
전라도 거주	-2.911 (0.233)	-2.654 (0.274)	-1.828 (0.45)
제주도 거주	3.809 (0.484)	3.85 (0.479)	3.969 (0.461)
성별(남성=1)	-4.743*** (0.000)	-3.442 [†] (0.087)	-3.547 [†] (0.075)
교육년수	0.366 (0.552)	0.322 (0.599)	0.295 (0.625)
종교여부(신자=1)	-3.277** (0.005)	-4.303* (0.014)	-1.686 (0.379)
취업여부(취업=1)	-2.991* (0.029)	-1.145 (0.539)	1.001 (0.621)
월간가계소득	-3.39*** (0.001)	-3.186** (0.001)	-3.129** (0.001)
계급상향인식	-3.542*** (0.000)	-3.402*** (0.000)	-3.348*** (0.000)
가부장적가치관	-4.714*** (0.000)	-6.523*** (0.000)	-6.153*** (0.000)
우파권위주의성향	-1.804** (0.002)	-1.844** (0.002)	-1.945** (0.001)
연도더미(2016년)	2.629 [†] (0.053)	2.598 [†] (0.054)	8.323*** (0.000)
남성×취업자		-3.72 (0.107)	-3.641 (0.111)
남성×가부장적가치관		3.058* (0.012)	2.758* (0.023)
남성×종교신자		1.649 (0.478)	1.712 (0.460)
2016년×취업자			-5.399* (0.026)
2016년×가부장적가치관			-0.179 (0.881)
2016년×종교신자			-7.039** (0.003)
adjusted R ²	0.2312	0.2415	0.2576
predictive MSE	203.6696	200.0259	201.0661

[도표 18] 최소자승추정(OLS) 회귀분석표

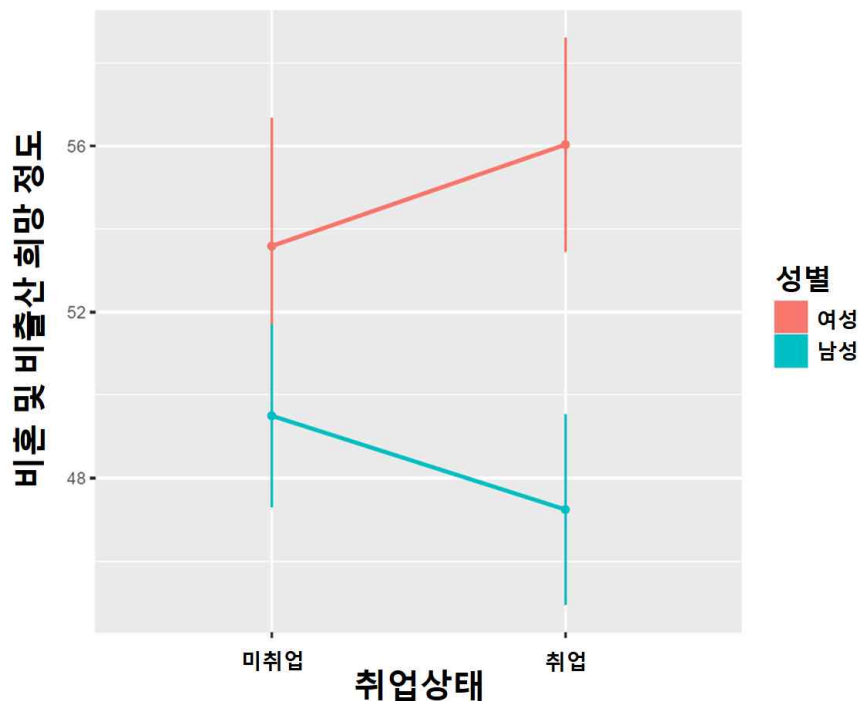
adaptive LASSO	Model #1	Model #2	Model #3
Intercept	56.140028	54.321055	52.7511689
연령	1.984011	2.033640	2.2457418
경기도 거주	×	×	-0.3611691
강원도 거주	-4.197926	-3.429962	-3.3558530
충청도 거주	-2.505070	-1.921567	-1.6683311
경상도 거주	×	×	0.2413379
전라도 거주	-2.509314	-1.873148	-1.2224402
제주도 거주	2.919783	3.295763	3.4658308
성별(남성=1)	-4.507567	-1.973757	-2.8768741
교육년수	×	×	×
종교여부(신자=1)	-3.047376	-3.150077	-0.3111174
취업여부(취업=1)	-2.593279	×	×
월간가계소득	-2.734920	-2.293278	-2.4946935
계급상향인식	-2.909791	-2.538913	-2.7990988
가부장적가치관	-4.779172	-6.157883	-6.0116792
우파권위주의성향	-1.692458	-1.730276	-1.7854434
연도더미(2016년)	2.273953	2.079063	7.5536992
남성×취업자		-4.482833	-2.9606839
남성×가부장적가치관		2.355960	2.2383044
남성×종교신자		×	×
2016년×취업자			-4.2417965
2016년×가부장적가치관			×
2016년×종교신자			-7.1726592
minimum λ	0.1911794	0.2703325	0.1925423
predictive MSE	200.8783	199.52345	200.2056

[도표 19] 조정(adaptive) LASSO 회귀분석표

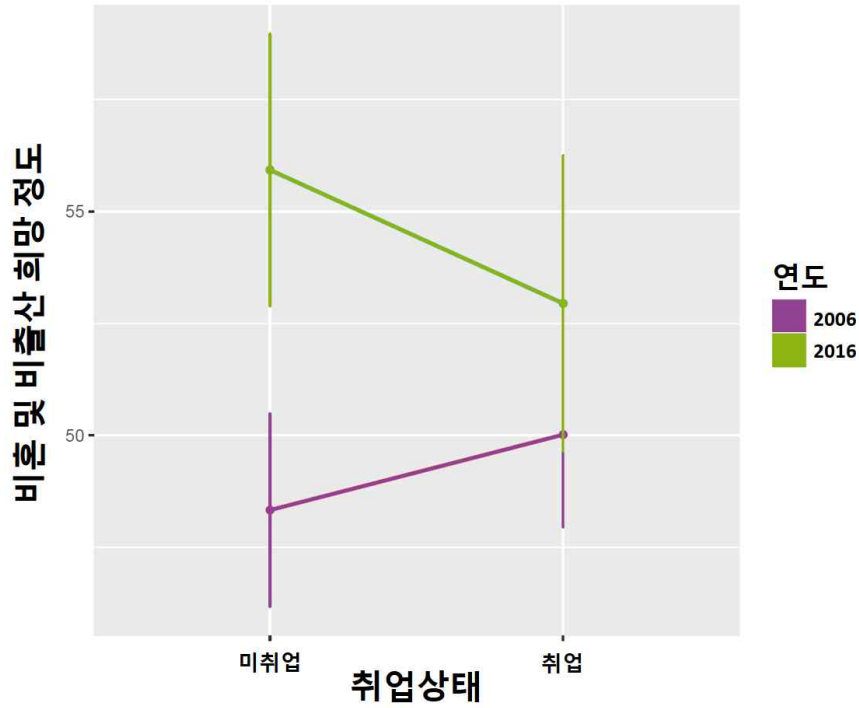
위의 [도표 18]과 [도표 19]는 각각 OLS와 조정 LASSO 회귀분석을 실시한 결과표이다. [도표 18]에서 관례적 유의미성은 회귀 계수의 위첨자에 표시했다. LASSO의 경우 p-value 대신 계수 축소(shrinkage)를 통해 변수를 선택하며 설명력이 약한 변수는 계수 크기가 0으로 축소된다. 계수가 0으로 축소되어 사실 상 제거된 변수는 표에서 × 표시로 처리하였다. 회귀계수에 관한 이하의 모든 설명은 최종 모델인 Model 3을 기준으로 한다.

분석 결과 OLS에서는 연령과 성별, 월간가계소득, 계급상향인식, 가부장적가치관, 우파권위주의성향, 연도더미가 유의미한 변수로 선택되었으며 교차 항에서는 남성×가부장적가치관과 2016년×취업여부, 2016년×종교신자여부가 선택되었다. (p-value 0.1 이하 기준) 반면 조정 LASSO 회귀분석의 경우 지역더미와 종교신자여부, 남성×취업여부의 세 개 변수가 추가적으로 선택되었다.

모델 및 기법 성능을 비교하기 위해 cross validation을 통해 전체 데이터의 55%를 training set으로 활용, 나머지 45%의 test set에 대해 예측을 실시한 결과, 조정 LASSO 모델을 사용 시 예측평균제곱오차(prediction mean squared error)가 201.0661에서 200.2056으로 감소했다. 최종 LASSO 회귀모델 기준으로 독립변수가 한 단위(1SD) 증가할 때마다의 비혼·비출산·탈가족주의 가치관의 변화를 설명해보자면, 나이가 많을수록 2.17%pt, 여성일수록 2.8%pt, 종교를 믿지 않을수록 0.11%pt, 소득이 적을수록 2.28%pt, 자신이 실제보다 잘 산다고 착각할수록 2.58%pt, 가부장적 가치관과 우파 권위주의의 성향이 덜할수록 각각 5.88%pt와 1.7%pt 비혼·비출산 경향이 증가하는 것으로 나타났다. 이하는 교차효과항들의 분석 내용이다. 도표에 표시된 에러 바는 95% 신뢰구간(confidence interval)을 의미한다.



[도표 20] 취업상태와 성별의 교차효과표



[도표 21] 취업상태와 연도의 교차효과표

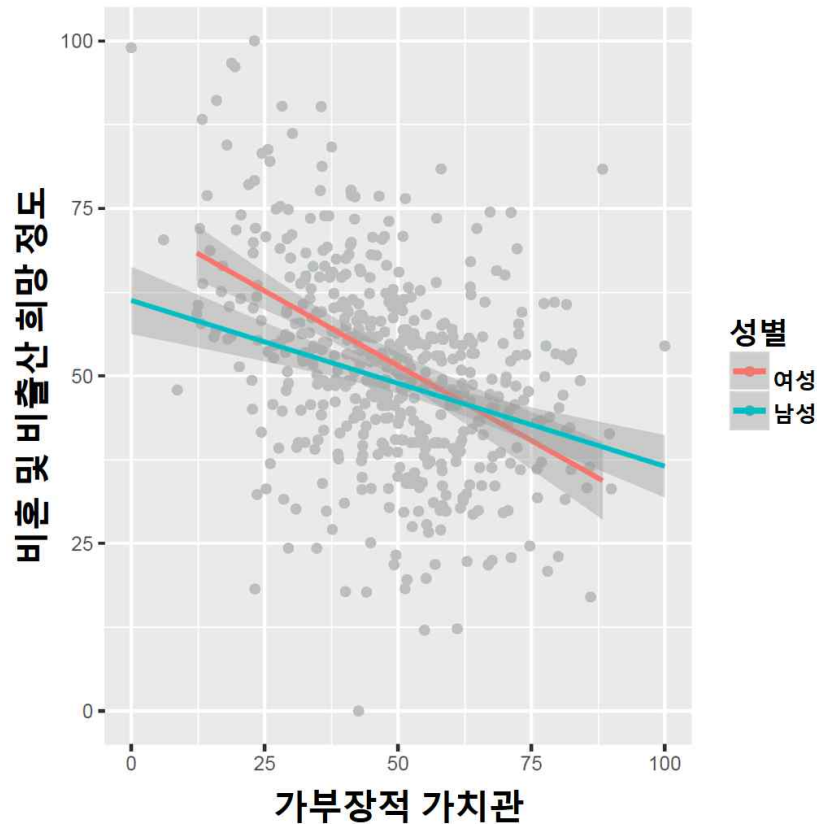
[도표 20]과 [도표 21]을 통해 알 수 있듯 전반적으로 여성은 남성에 비해 비혼·비출산 경향이 강하며, 이러한 차이는 취업 이후에 확대된다. 비혼·비출산 경향이 감소하는 남성과 달리 여성은 오히려 강화되기 때문이다. 연도를 기준으로 살펴보자면, 2006년에는 취업자의 경우 미취업자에 비해 비혼·비출산 희망 정도가 더 높았으나 2016년에는 경향이 역전되었다.



[도표 22] 취업상태와 성별의 교차효과표 연도 간 비교

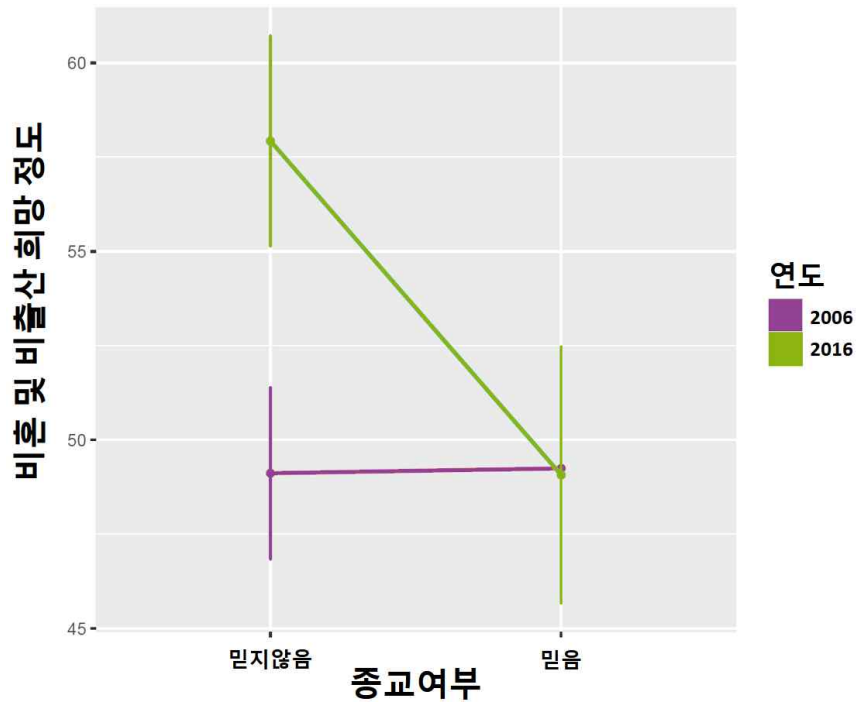
취업상태와 성별 간 교차효과표를 연도에 따라 나누어보면 좀 더 명확한 정보를 얻을 수 있다. 2006년의 여성을 제외하고는 모두 우하향하는, 즉 취업자가 비혼·비출산 경향이 더 감

소하는 양상을 보이고 있기 때문이다. 이는 다시 말해 지난 10년간의 사회구조적 변화가 남성이 아닌 여성의 혼인·출산관에 더 많은 영향을 끼쳤음을 시사한다. 그리고 바로 이 지점에서 선행연구에서 언급한 여성의 사회진출과 혼인·출산의 문제로 돌아가 보자면, 위의 [도표 22]는 한국 사회가 여성의 노동시장 참여 임계점을 넘겼고 이제는 직업적 생활과 가정 생활이 양립 가능성을 암시하기도 한다. 다만 이를 무조건적으로 긍정적으로만 볼 수 없는 것은, 이러한 인식의 전환이 여성의 노동 시장에서의 자녀 양육 및 결혼 생활의 병행 여건 개선의 결과인지, 아니면 경제적 상황의 절대적인 악화에 따라 가계 이익 극대화를 위해 여성들이 택한 전략적 선택인지 그 인과적 관계를 확인할 수 없기 때문이다.



[도표 23] 가부장적 가치관과 성별의 교차효과표

[도표 23]은 가부장적 가치관의 성별에 따른 조절효과를 보여준다. 가부장적 경향이 약할 경우 여성이 남성보다 비혼·저출산 희망 정도가 더 높지만 가부장적 가치관 점수가 약 60점(전체 100점 만점)을 넘어설 경우 그 양상이 역전되는 것을 볼 수 있다. 층위가 다르기에 본 연구의 미시적인 분석을 거시적 상황에 그대로 적용할 수는 없겠지만, 만약 가부장적 가치관이 해체되어가는 현 추세는 계속되고, 또 개별 행위자들이 이러한 추세에 직접적인 영향을 받는다는 느슨한 가정을 전제할 경우 우리는 남아 선택을 통해 1970-1990년대에 희소해진 여성이 결혼 시장의 참여에 더욱 더 소극적으로 변화함에 따라 비혼·저출산 현상이 더욱 심화될 것이라고 예상해 볼 수 있다.



[도표 24] 종교 여부와 연도 간의 교차효과표

마지막으로 [도표 24]는 종교 신자 여부와 연도 간의 교차 효과를 보여준다. 종교 신자들은 10년의 기간 동안 가족관이 크게 변화하지 않았지만, 비신자들의 경우 약 10%pt 가량 급격히 탈가족주의 성향이 높아졌음을 알 수 있다. 최근 계속되고 있는 탈세속화, 즉 비(非)종교 신자 인구의 증가 추세를 감안할 때, 상대적으로 전통적 가족관을 고수해 결혼과 출산에 적극적인 집단은 향후에도 감소할 것이라고 판단할 수 있다.

5. 결론 및 함의

연구자는 한국종합사회조사(KGSS)의 2006년도와 2016년도 데이터를 바탕으로 한국 사회의 비혼·저출산·탈가족주의 가치관이 인구통계적·사회심리적 요인에 어떠한 영향을 받는지, 그리고 성별과 시기의 교차 효과에 따라서 어떻게 상이하게 나타나는지를 조정 LASSO 회귀 분석을 통해 알아보는 것을 목표로 하였다. 본 연구는 크게 두 가지 점에서 의미를 지닌다.

첫 번째, 미시적인 행위자들의 심리적·경제적·성차(性差)적 요인이 결혼·출산·가족관과 어떠한 관계를 맺는지를 10년의 시차를 둔 데이터를 통해 다루고 있다. 그 결과 탈가부장적 가치관, 탈권위주의적 정치 성향이 비혼·저출산과 유의미한 관계를 맺는다는 직관적인 사실을 최신의 데이터를 바탕으로 재확인 할 수 있었다. 또한 이를 통해 구조적으로 저출산 가치관이 향후에도 계속 확산될 것이라는 다소 부정적인 예측 역시 가능했다.

두 번째, 교차효과를 활용한 조절 모형들은 시기와 성별에 따라 어떻게 혼인·출산관이 상이하게 나타나는지를 보여준다. 특히 OLS와 달리 LASSO 정규화 기법을 활용함으로써 노동 시장에 참여한 여성들과 그렇지 않은 여성들 간의 차이를 유의미하게 확인할 수 있었다. 기존의 국내 연구자들 중 상당수는 한국에서 여전히 일과 가정이 양립 불가능하고, 또 노동시장에의 참여는 혼인과 출산을 저해하는 요인으로 기능한다고 주장해왔다. 그러나 2006년과

2016년 양자 간 데이터를 비교해본 결과, 2006년에는 해당 주장들에 부합하는 결과를 관찰할 수 있었지만 2016년에 들어서는 취업한 여성들이 취업한 남성과 마찬가지로 미취업자들에 비해 더 높은 결혼·출산 경향을 보이는 것으로 나타났다. 물론 앞서 언급한 바와 같이, 이를 긍정적인 신호로만 해석할 수는 없다. 가령 미국과 일본에서는 노동 시장의 유연성 확대에 따른 중년 여성들의 재진입을 통해 긍정적이지 않은 방향으로 일과 가정의 양립이 실현되었기 때문이다. 이는 실질적인 일과 가족의 양립을 위해 제도적 지원을 확대한 결과로 해당 정책적 목적을 달성한 북유럽 사민주의 국가들과는 다른 양상이라고 할 수 있다. (신경아, 2010) 따라서 본 연구에서 확인된 ‘여성 노동시장의 임계점’을 지난 한국 사회가 어떻게 일과 가정의 양립을 지원해야 긍정적인 방식으로 혼인·출산의 확대를 추구할 수 있을지에 대해 향후 정책적인 논의가 필요할 것이다.

6. 참고문헌

자료출처 : 김상욱. “한국종합사회조사”. 자료산출기관: 성균관대학교 동아시아학술원 서베이 리서치센터. 자료제공기관: 한국사회과학자료원.

김경근, 최윤진. "교육열 현상으로서의 저출산." *교육사회학연구* 27.2 (2017): 1-34.

김영미, and 계봉오. "이행의 계곡에서 무슨 일이 벌어지나." *한국여성학* 31.3 (2015): 1-30.

김영미. "출산과 성평등주의 다층분석." *경제와사회* 112 (2017): 41-74.

계봉오, and 김영미. "한국여성의 결혼, 출산 및 경제활동 참가 형태에 대한 연구." *사회연구* (2014): 79-113.

서동희. "저출산 극복의 새로운 길, 저출산 적응정책." *공공사회연구* 5.1 (2015): 69-102.

신경아. "저출산대책의 쟁점과 딜레마." *페미니즘 연구* 10.1 (2010): 89-122.

엄동욱. "우리나라는 저출산함정에 빠진 것인가?*-저출산함정 가설의 검증과 함의." *한국인구학* 32.2 (2009): 141-159.

이상은, and 김희찬. "한국의 낮은 아동빈곤과 저출산의 역설, 그리고 정부 가족지출." *한국사회복지정책학회 춘계학술대회자료집 2017* (2017): 261-278.

정성호. "저출산의 사회경제적 요인과 향후과제." *사회과학연구* 48.2 (2009): 1-22.

Poston Jr, Dudley L., Heather Terrell Kincannon, and Jungwon Yoon. "Socioeconomic development and fertility in the People's Republic of China and the Republic of Korea." *Ultra-low Fertility in Pacific Asia, Trends, Causes, and*

Policy Issues (2009).

Larsen, Ulla, Woojin Chung, and Monica Das Gupta. "Fertility and son preference in Korea." *Population Studies* 52.3 (1998): 317–325.

Suzuki, Toru. "Korea's Strong Familism and Lowest-Low Fertility." *International Journal of Japanese Sociology* 17.1 (2008): 30–41.

Park, Chai Bin, and Nam-Hoon Cho. "Consequences of son preference in a low-fertility society: imbalance of the sex ratio at birth in Korea." *Population and development review* (1995): 59–84.

Sheps, Mindel C. "Effects on family size and sex ratio of preferences regarding the sex of children." *Population Studies* 17.1 (1963): 66–72.

Jones, Gavin W., and Bina Gubhaju. "Factors influencing changes in mean age at first marriage and proportions never marrying in the low-fertility countries of East and Southeast Asia." *Asian Population Studies* 5.3 (2009): 237–265.

McDonald, Peter. "Explanations of low fertility in East Asia: A comparative perspective." *Ultra-low fertility in Pacific Asia: Trends, causes and policy issues*. Routledge, Taylor & Francis Group, 2009.

Anderson, Thomas, and Hans-Peter Kohler. "Education fever and the East Asian fertility puzzle: A case study of low fertility in South Korea." *Asian population studies* 9.2 (2013): 196–215.

Jones, Gavin W., Paulin Tay Straughan, and Angelique Chan. "Very low fertility in Pacific Asian countries." *Ultra-low fertility in Pacific Asia: Trends, causes and policy issues* (2008): 1–22.

Kang, Hye Jung, and Hyo Jin Song. "Recognition and Response to Low Fertility—Policy Response to Low Fertility in Japan." *한국위기관리논집* 13.2 (2017): 1–16.

Lutz, Wolfgang, Vegard Skirbekk, and Maria Rita Testa. "The low-fertility trap hypothesis: Forces that may lead to further postponement and fewer births in Europe." *Vienna yearbook of population research* (2006): 167–192.

Rindfuss, Ronald R., Karen Benjamin Guzzo, and S. Philip Morgan. "The

changing institutional context of low fertility." *Population Research and Policy Review* 22.5 (2003): 411–438.

Morgan, S. Philip. "Is low fertility a twenty-first-century demographic crisis?." *Demography* 40.4 (2003): 589–603.

Wasserstein, Ronald L., and Nicole A. Lazar. "The ASA's statement on p-values: context, process, and purpose." *The American Statistician* (2016).

Trafimow, D. and Marks, M. *Basic Appl. Soc. Psych.* 37, 1-2 (2015).

Open Science Collaboration. "Estimating the reproducibility of psychological science." *Science* 349.6251 (2015): aac4716.

Tibshirani, Robert. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* (1996): 267–288.

Zou, Hui. "The adaptive lasso and its oracle properties." *Journal of the American statistical association* 101.476 (2006): 1418–1429.

Hoerl, Arthur E., and Robert W. Kennard. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12.1 (1970): 55–67.