

빅데이터 기반의 출산율 변동 예측

주세민* · 옥성환** · 황경태***

요약

본 연구에서는 육아에 대한 공포 등 심리적 요인이 출산율에 미치는 영향을 실증적으로 분석하였다. 2000년~2018년까지 육아에 대한 부정적인 기사가 전체 사회 기사에서 차지하는 비중을 바탕으로 지표를 산출하였다. 지표 분석 결과, 지표가 증가하면 3년 뒤의 출산율은 떨어지는 것으로 나타났다. 이러한 결과는 상관관계 분석, 단순 회귀분석, VAR 분석에서도 일관적으로 나타났다. 그랜저 인과관계 분석 결과, 지표와 3년 뒤 출산율의 관계는 단순 상관관계가 아닌 인과관계에 있음을 알 수 있었다. 연령대별로도 차이를 보였는데, 20~30대 여성의 출산율은 지표에 유의한 반응을 보였으나, 40대 출산율은 반응을 보이지 않았다. 또한 지표가 상승하면 1아 출산율에는 영향을 미치지만, 2아/3아 이상의 출산율에는 영향을 미치지 않는 것으로 나타났다. 이것은 여성의 나이가 어릴수록 육아에 대한 부정적인 기사에 영향을 많이 받지만, 이미 출산/육아를 경험해본 사람들에게는 큰 영향을 미치지 못한다는 직관과도 일치한다. 본 연구는 뉴스 빅데이터를 단순한 키워드 언급량 변화 모니터링이라는 한정된 용도를 벗어나, 사회 현상을 예측하는데 유의미한 지표를 추출해 냈다는데 의미가 있다. 또한 이러한 빅데이터 기반의 지표는 출산율에 대한 3년의 선행성이 있기 때문에 미리 감지할 수 정보를 제공한다는 장점이 있다.

주제어 : 뉴스, 빅데이터, 불확실성, 출산율, 심리적 요인, 부정적 기사

Forecasting Birthrate Change based on Big Data

Joo Se-Min* · Ok Seong-Hwan** · Hwang Kyung-Tae***

Abstract

We empirically analyze the effects of psychological factors, such as the fear of parenting, on fertility rates. An index is calculated based on the share of negative news articles on child care in all social articles from 2000 to 2018. The analysis result shows that as the index increases, the fertility rate after three years falls. This result is repeated in the correlation analysis, simple regression, and VAR analysis. According to Granger causality analysis, it is found that the relation between the index and the fertility rate after three years is not just a simple correlation but a causal relationship. There are differences among age groups. The fertility rate of women in their 20s and 30s shows a significant response to the index, but that of the 40s does not. The index affects the birthrate of first child, but do not affect the birthrate of second or more children. These results are consistent with the intuition that younger women are more likely to be affected by the negative articles about parenting, but not to those who have already experienced childbirth. This study is meaningful in that a significant index for predicting social phenomena is extracted beyond the limited use of news big data such as a simple keyword mention volume monitoring. Also, this big data-based index is a 3-year leading indicator for fertility, which provides the advantage of providing information that helps early detection.

Keywords : news, big data, uncertainty, birth rate, psychological factors, negative news

Received Sep 24, 2019; Revised Oct 10, 2019; Accepted Oct 15, 2019

* First author, Ph.D Economics, Sogang University(sewoogi@naver.com)

** Author, Ph.D Candidate, Dongguk University(oakhill@naver.com)

*** Corresponding author, Professor, Dongguk University (kthwang@dongguk.edu)

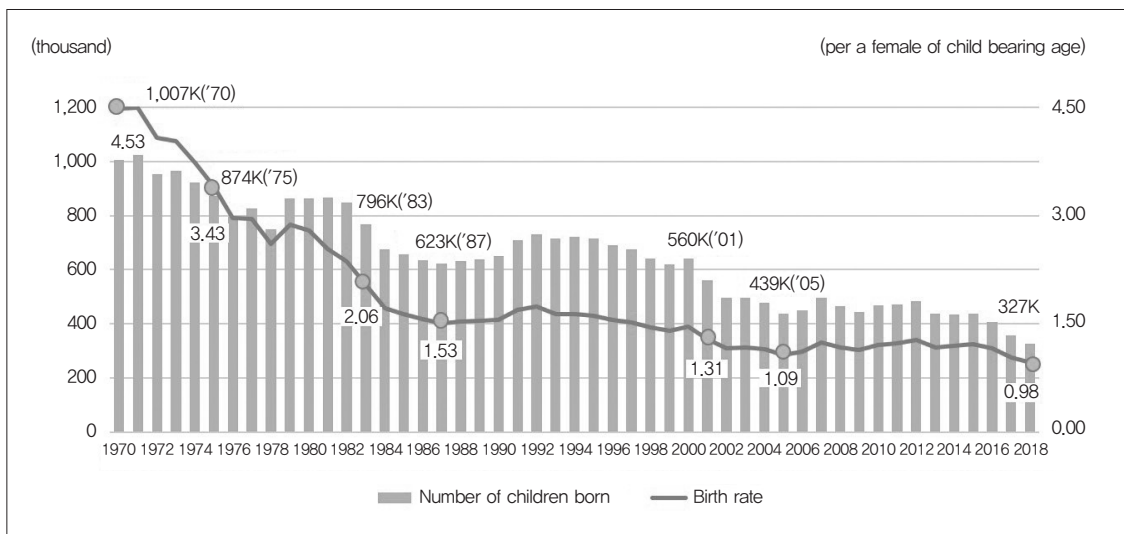
I. 서론

최근 한국의 출산율은 지속적으로 낮아지고 있다고 알려져 있다(〈그림 1〉 참조). 하지만 지속적으로 낮아진 것은 아니고, 2005년에 합계 출산율이 1.08로 최저치를 기록한 이후 2013년에는 1.30까지 증가했다가 2018년에 다시 0.98로 떨어졌다(Statistics Korea, 2018). 이를 통해서 출산율의 시계열 변동을 일으키는 요소가 있을 수 있다고 추론해 볼 수 있다.

다양한 요인들이 출산율에 영향을 미치는 것으로 보고되고 있다. 출산율에 대한 전통적인 연구에서는 주로 사회경제적인 요인과 인구사회적인 요인을 제시하고 있다. Lovenheim(2013)에 의하면, 가정의 소득 수준이 높으면 출산율이 높아진다고 한다. 출산 비용은 출산 의향에 영향을 미친다는 연구 결과들이 있다(Choi, et al., 2014; Shin, 2008). 그리고 직장여성의 출산으로 인한 경력 단절은 출산율을 5~10% 낮추는 효과가 있다고 한다(Del Bono, 2012). 또한 평균 수명의 증가가 출산율 감소를 설명한다는 연구도 있다

(Angeles, 2010). 1900년 이후 여성의 역할 증대와 인구구조 변화가 가속화 되면서, 이러한 사회경제적 요인과 인구사회적인 요인들은 거시적인 측면에서 활발히 연구되고 있다.

다른 한편으로는 가치관과 같은 사회심리적인 요인으로 출산율을 설명하는 연구들이 있다. 출산에 영향을 미치는 가치관에는 “결혼관”, “자녀관”, “성역할관” 등이 있다. 결혼은 당연한 것이 아니라 선택이라는 결혼관이 출산에 영향을 미친다는 연구들이 있다(Kim, et al., 2002; Kim, et al., 2003; Lee, 2003). Lee(2006)은 전국 결혼 및 출산 동향조사(설문) 결과를 토대로, 결혼에 대한 부정적인 태도가 최근에 높아졌으며, 이것이 출산에 부정적인 영향을 미친다는 연구 결과를 제시하였다. 자녀관의 변화와 출산율의 변동을 연구한 Ministry of Women(2001)는 일정 규모 이상의 자녀를 출산하는 것을 자연스럽게 받아들이는 현상이 약해지고 있다고 한다. 한편 여성은 육아와 가사노동을 맡아야 한다는 전통적인 성역할관이 출산율에 영향을 미칠 수 있다. Künzler(2002)에 의하면, 가



Source: 2018 Birth Statistics (Finalized), Statistics Korea.

〈그림 1〉 출산율 추이
〈Fig. 1〉 Birth Statistics

사와 육아의 부담이 평등한 국가일수록 출산력이 높다고 한다.

각종 출산장려 정책의 효과를 분석한 연구들도 있다. Hong, et al.(2013)은 보육비 지원 정책이 추가 출산에 미치는 영향을 분석하였는데, 분석 결과 보육비 지원은 추가 출산이나 출산 의향을 약 2% 증가시키는 효과가 있다고 한다. Lee and Myung(2015)의 연구에서는 임신/출산지원, 신생아 의료지원, 보육비용 지원, 일/가정 양립제도(산전후 휴가, 육아휴직 등) 등의 네 가지 요인이 출산 의향에 미치는 영향을 조사하였는데, 지방정부의 보육지원정책만이 출산결정에 영향을 미치는 것으로 나타났다. 집행된 정책과 실제 출생율 간의 관계를 살펴본 연구도 있는데, Langridge, et al.(2012)은 출산 장려금이 10대 또는 사회 취약계층의 출산율만 늘린다는 것을 시계열 데이터를 사용하여 입증하였다. Hong, et al.(2016)은 지역구마다 출산 장려금에 차이가 있으며, 출산 장려금이 높은 지역일수록 출산율이 높은 것으로 조사되었다.

이러한 연구 결과를 바탕으로 출산율에 영향을 미치는 요인들을 다음의 <표 1>과 같이 정리할 수 있다.

본 연구에서는 사회 심리적인 영향으로 출산과 육아에 대한 부담/공포가 출산율에 미치는 영향에 대해서 연구하고자 한다. 사회 심리적인 요인에 주목하는 첫 번째 이유는 최근의 출산율 하락이 각종 출산 장려 정책의 효과로 설명되지 않는 부분이 크기 때문이다.

<표 2>에서 알 수 있듯이, 2012년 이후에 보편적 무상 복지의 일환으로 전국민 무상 보육이 실시되었고, 각종 출산과 육아를 직간접적으로 지원하는 정책이 실행되면서 실질적인 육아환경은 개선되었다고 볼 수 있다. 하지만, 본격적인 보편적 무상 복지가 실현된 2012년 이후 출산율은 2012년 당시 출산율(1.297)보다 높아지지 않았다(2016년 출산율은 1.172) (Statistics Korea, 2018). 즉, 실질적인 육아 환경 개선이 출산율을 높이는데 영향을 미쳤다고 말하기 어렵다고 판단된다. 사회 심리적인 요인에 주목하는 두 번째 이유는 최근에 다양한 사회현상을 설명하는데 있어서 심리적인 불확실성이 유효한 원인이라는 연구결과가 나오고 있기 때문이다 (Jurado, et al., 2015 : Nimark, 2014 : Baker, et al., 2016).

본 연구에서 분석하는 출산과 육아에 대한 부담/공포는 기존에 연구되던 자녀관과는 다른 개념이다. 자녀관은 자녀가 필요한지, 필요하다면 얼마나 필요한지 등에 대한 가치관의 영역이라면, 출산과 육아에 대한 부담/공포는 향후 본인에게 일어날지도 모르는 육체적, 금전적 어려움에 대한 심리적인 불확실성의 영역이다.

실물경제가 변하지 않는데도 불구하고 미래에 대한 불확실성이라는 심리적인 요인으로 경제주체들의 행동이 변한다는 다양한 선행 연구들이 있다. 이 때 불확실성을 어떤 데이터로 정의할 것이냐가 문제되는데,

<표 1> 출산율에 영향을 미치는 요인
<Table 1> Factors Affecting Birth Rate

Factor	Elements
Social/Economic	opportunity cost for childcare, educational level of women, types of occupation, family income level
Population/Social	life expectancy, age of women at the first marriage, population of women at the childbearing age, survival of parents
Childbirth Promotion Policies	1) direct financial support policy - childbirth grant, childcare subsidy 2) indirect support policy - maternity leave benefit, medical expense discount
Social/Psychological	values (on marriage, child, gender), fear/burden of childbirth/child care

〈표 2〉 양육지원 정책
 〈Table 2〉 Child Support Policies

(단위: 천원) (unit: thousand won)

		2009~2011	2012	2013	2017
child support grant	0~2	[bottom 70%] 83~383	[all] 288~394	[all] 288~394	[all] 313~430
	2~4	[bottom 70%] 51~191	[bottom 70%] 220	[all] 220	[all] 220
	5	-	[all] 220	[all] 220	[all] 220
child benefit		[next higher] 100~200	[next higher] 100~200	[all] 100~200	[all] 100~200
remark		free care from July 2009	free care for all groups	free care for 5-year old & below	maintain same benefit level

전통적으로 사용되는 지표는 변동성 지수(Volatility Index: VIX)이다. 이는 시계열 변수의 분산으로 나타내며, 일정 기간 시계열 데이터의 분산이 커지면 경제 주체는 불안감을 느끼며 보다 큰 리스크 프리미엄을 기대하게 된다는 것이다. 그러나, 최근에는 이러한 전통적인 변동성 지수와는 다른 방법으로 불확실성을 포착하려는 시도가 있다. Jurado, et al.(2015)은 VIX와 같이 특정 지표의 증감 또는 분산에 변동이 생긴다고 경제학적 의사결정이 영향을 받는 것은 아니라고 하였다. 즉, 변동의 발생 여부가 아니라, 경제 예측이 어려워진 정도가 불확실성 판단의 기준이 되어야 한다는 것이다. 따라서, Jurado가 정의하는 불확실성이란 현재 가용한 데이터로 종속변수를 예측할 때 예측 불가능한 성분의 변동폭이다.

한편, 최근 빅데이터 기술이 발전하면서 사람들이 접하는 뉴스에서 불확실성을 포착하려는 시도도 있다. Nimark(2014)은 “개가 사람을 물었다”는 뉴스는 일반적으로 발생할 수 있는 뉴스이지만, 반대로 “사람

이 개를 물었다”와 같이 일반적이지 않은 뉴스가 불확실성을 야기한다고 한다. 이렇게 일반적이지 않은 뉴스에는 경제 주체들이 똑같이 반응하지 않으며, 이러한 현상을 불확실성이 높은 현상이라고 판단하였다.¹⁾ Nimark(2014)과 유사하게 Baker, et al.(2016)도 빅데이터를 통하여 불확실성을 설명하고자 하였다. 다만, Nimark와 다르게 Baker, et al.(2016)는 특정 단어가 포함되어 있는 뉴스가 많아질수록 경제 주체들은 불확실성을 크게 느낀다고 하였다. 이 연구는 불확실성이라는 다소 모호한 개념을 빅데이터 기술을 사용하여 정량화하였으며, 실제 경제 예측에 설명력을 가짐을 입증함으로써 미국뿐만 아니라 다른 국가에서도 이와 유사한 연구가 이루어지고 있다(Arbatli, et al., 2017). Nimark(2014) 방법론을 활용하였을 경우, 일반적이지 않은 뉴스를 접하고 반응하는 구성원은 특정 경제 주체에 한정될 우려가 있으며, 전체 구성원들이 어떻게 반응하는지를 고찰하기 쉽지 않을 것으로 판단하였다. 반면 Baker, et al.(2016)의 방법

1) 예를 들어, “개가 사람을 물었다.”라는 뉴스에는 모두 비슷한 반응을 보이지만, “사람이 개를 물었다.”라는 일반적이지 않은 뉴스에는 사람들마다 다르게 반응함으로써 전체 경제의 불확실성이 커진다는 것이다.

론은 특정 주제에 대해 기사가 쏟아질 경우 전체 구성원들이 어떻게 반응하는지 바라보기 용이하다는 장점이 있어 출산율 변동 예측에는 보다 적합할 수 있다고 판단하였다.

본 연구에서는 객관적이고 거시적인 데이터로서 빅데이터를 활용하여 출산율의 변동을 예측하고자 한다. 본 연구의 주요한 의의는 다음과 같이 세 가지로 정리할 수 있다.

첫째, 빅데이터를 활용하여 출산율이라는 사회 현상을 설명하려고 시도하는 점이다. 이 분야의 기존 연구들은 대부분 가임기의 여성들을 대상으로 한 설문조사 데이터에 의존하고 있다. 하지만, 최근 들어 설문에 의존하지 않고 빅데이터를 기반으로 다양한 불확실성을 포착하는 연구의 진보가 이루어졌다. 앞서 언급한 Baker, et al.(2016)는 빅데이터 기법을 활용하여 불확실성 지표를 정의하였는데, 뉴스 빅데이터로부터 경제 현상을 예측할 수 있는 의미 있는 지표를 도출했다. 이 연구에서는 일간지에 언급된 단어의 빈도를 기반으로 경제정책 불확실성(Economic Policy Uncertainty: EPU) 지표를 개발하였는데, 이 지표는 걸프전, 911 테러, 리만브라더 사태, 2011년 부채 제한 논쟁 (Debt Ceiling Dispute) 등을 잘 설명한다. 또한 기본 EPU 지표 외에도 조세, 재정정책, 국가 보안, 헬스케어 등 11개 유형의 추가적인 지표도 개발하였다. 여기에 영향을 받아 미국 이외의 국가에서도 이러한 방법론을 활용한 연구가 활발히 진행되고 있다. Arbatli, et al.(2017)은 EPU 지표 산출 방법을 활용하여 일본의 EPU 지표를 개발하였다. 본 연구에서는 기준을 명확히 세울 수 있으며, 다른 국가에서 검증된 Baker, et al.(2016)의 방법론을 분석 방법으로 활용한다.

둘째, “출산에 대한 공포”, “육아에 대한 막연한 두려움” 등 기존의 “결혼관”, “자녀관”, “성역할관” 등으로는 규명할 수 없는 사회 심리적 요인을 분석하는 점

이다. 기존의 출산율에 대한 연구는 주로 “여성의 사회 진출”, “가정 소득”, “출산 및 육아 지원 정책” 등 실물 경제가 변하는 경우 그에 대한 출산율의 변동을 설명하고 있다. 그러나 심리적 부담이나 공포가 우리나라의 출산율 변동에 미치는 영향을 분석한 연구는 아직까지 본격적으로 논의되고 있지 않다. 최근 “헬조선”, “독박육아”, “육아 비용 증가” 등 부정적인 기사와 인터넷의 내용들을 접하면서 젊은 세대들이 실제로 다 훨씬 출산과 육아에 두려움을 가질 가능성이 높다. 이에 따라 본 연구에서는 Baker, et al.(2016)의 방법론을 이용하여 사회 심리적 요인이 출산율에 미치는 영향을 알아보고자 한다.

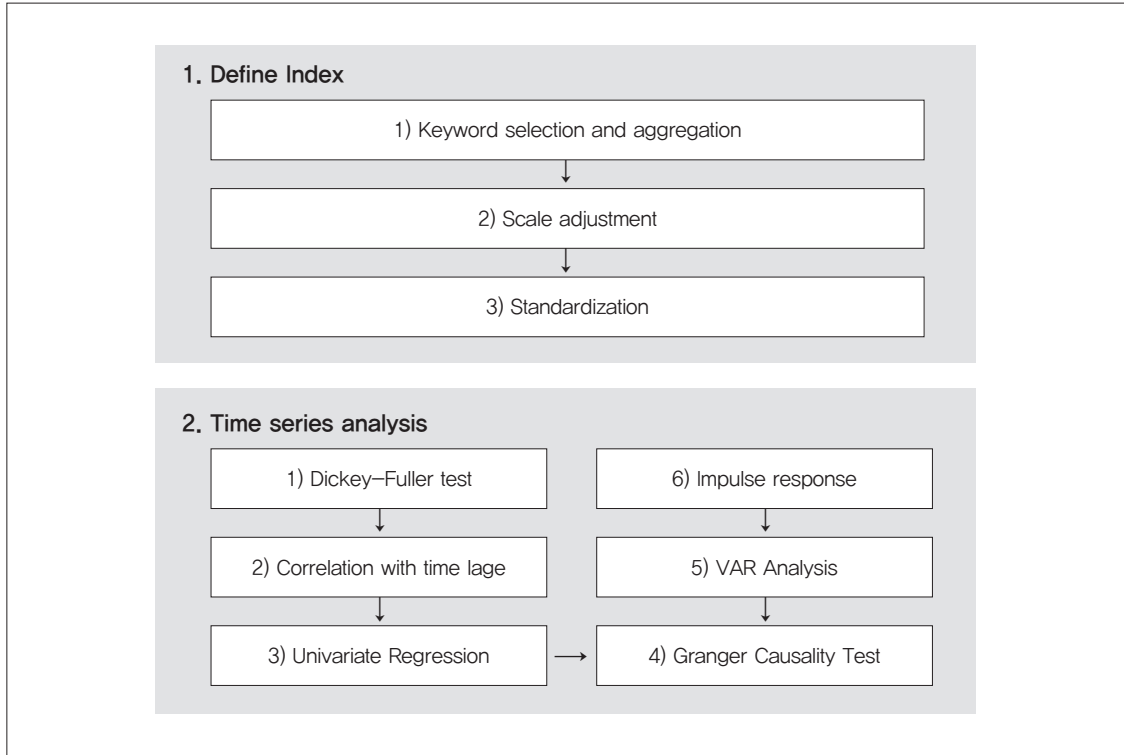
셋째, 뉴스나 소셜미디어와 같은 텍스트 기반 빅데이터의 활용성을 제고할 수 있는 방향성을 제시할 수 있을 것으로 기대된다. 우리나라의 뉴스 빅데이터 활용 현황은 키워드 기반의 주제 분류, 연관관계 분석 등이 주를 이룬다.²⁾ 미국의 경우, 연방준비제도경제데이터(Federal Reserve Economic Data: FRED) 사이트를 통하여 앞서 언급한 EPU 지표를 생성하여 제공하고, 실물 경제 예측에 활용하고 있다. 우리나라도 우리 실정에 맞는 빅데이터 기반의 다양한 지표 산출 및 경제 예측을 시도해 볼 필요가 있다고 생각된다.

II. 연구 방법론

본 장에서는 먼저, 본 연구에서 차용한 Baker, et al.(2016)의 방법론에 대해 간략하게 설명하고, 다음으로 본 연구의 세부 절차에 대해 설명한다.

Baker, et al.(2016)의 분석 1단계는 불확실성을 측정하기 위해 전체 신문 기사 중에서 특정 단어가 나타나는 기사의 수를 집계하여 지표를 만드는 것이다. 이 연구에서는 주요 일간지 10개의 기사를 수집하여 다음 세 가지 범주의 단어가 나타나는 기사의 수를 집계하였다. (1) 경제 관련 단어: “economic” 또는

2) 한국언론진흥재단 빅카인즈 서비스 및 소셜 빅데이터 분석 솔루션(링크 브릭스, D-Map, ODPia등) 검토 결과



〈그림 2〉 연구 절차

〈Fig. 2〉 Analysis Procedures

“economy”, (2) 불확실성 관련 단어: “uncertain” 또는 “uncertainty”, (3) 정책 관련 단어: “Congress”, “deficit”, “Federal Reserve”, “legislation”, “White House” 중 적어도 하나를 포함하는지 여부. 정책 관련 단어의 선정은 사람들이 직접 참여하는 감사(Audit) 프로세스에 따라 실제 기사를 읽으면서 해당 정책을 나타내는 단어를 선별하였다. 이러한 단어가 포함된 기사의 숫자가 집계되면, 추가적인 조정을 거친다. 해당 연도에 불확실성에 관련된 기사가 많아진 것이 아니라, 당해 기사 전체의 수가 많아졌을 수도 있기 때문에, 전체 기사의 양 대비 불확실성을 나타내는 기사의 수로 스케일 조정 (Scaling) 과정을 거친다. 그 후, 지표로 정의하기 위하여 특정 연도를 100으로 정의하는 표준화 (Normalization) 과정을 거친다.

2단계는 산출된 지표가 거시 지표에 미치는 영향을 시계열적인 분석 방법으로 분석하였다. 벡터자기회귀 (Vector Auto Regression: VAR) 모형을 이용하여 기사로부터 도출된 불확실성 지표가 실물 경제에 어떻게 반영되는지를 분석하였다.

이러한 기본적인 방법론을 본 연구에 적용한 절차는 〈그림 2〉에 정리되어 있고, 다음에서는 이에 대해 세부적으로 설명한다.

1. 지표 산출

출산 및 육아에 대한 부정적 기사를 기반으로 출산 및 육아 관련 불확실성 지표를 도출한다.

1) 키워드의 선별 및 집계

“육아”에 관한 “부정적인” 기사를 선별해내는 데는 두 가지 조건이 필요하다. 첫째는 “육아”에 대한 키워드이고, 둘째는 “부정적인” 키워드 이다. 본 연구에서는 주제를 “육아”로 한정하였기 때문에, 여기에 대한 키워드 선별은 특별히 진행하지 않았다.³⁾ 하지만, “부정적인” 기사를 검출해내는 키워드 선정은 Baker, et al.(2016)의 감사 프로세스를 준용하였다. 육아 관련 기사 200개를 선정하여 직접 읽으면서 육아에 대한 부정적인 기사들을 선별할 수 있는 키워드를 뽑아내었다.⁴⁾ 이에 따른, 기사 검색 조건은 다음과 같다. (1) “육아”, (2) “우려”, “걱정”, “불안”, “불만”, “독박”, “스트레스” 중 적어도 하나.

또한, Baker, et al.(2016)와 달리 일부 신문사에 한정하지 않고 네이버에서 검색 가능한 전체 기사 대상으로 1차 조사를 진행하고, 일간지만을 대상으로 2차 조사를 진행하였다.

2) 스케일 조정

시간의 흐름에 따라 전체 기사의 양이 달라지기 때문에, 연도별 기사의 수를 곧바로 활용하는 것은 적절하지 않다. 해당 기간의 사회 분야 기사의 총량으로 나누어 줌으로써 스케일을 맞춘다 (이 값을 라고 정의한다.)

3) 표준화

지표화하기 위해서 임의의 특정 기간의 의 평균값을 100으로 환산하고, 여기에 맞추어 전체 시계열 데이터를 표준화한다. 여기서는 2004년부터 2013년의 평

균값을 기준으로 하였다.⁵⁾ 즉, 2004년부터 2013년의 평균값을 $M = \frac{1}{2} \sum_{2004}^{2013} A_t$ 이라고 하면, 최종 지표는 아래와 같은 식으로 산출된다.

$$\text{지표} = A_t \frac{100}{M}$$

2. 시계열 분석

출산율은 매우 거시적인 데이터로서, 지표가 곧바로 출산율에 영향을 미칠 수는 없다. 아이를 가지기로 결심하고 출산하기까지 임신기간을 포함하여 적어도 1년의 시간이 필요하고, 출산에 대한 결심은 일반적으로 단시간에 이루어지지 않기 때문이다. 따라서 시계열 분석의 큰 방향은 (1) 연단위 데이터로 분석을 진행하고, (2) 지표의 시차(Lag)를 고려하는 것이다. 즉, 지표의 시차를 두면서 출산율과의 상관관계를 분석하고, 상관관계가 높은 경우에 출산율과 회귀 계수를 도출하여 유의성 검정을 진행한다(Hamilton, 1994).

시계열 데이터를 분석함에 있어서 확률적(Stochastic) 보행 과정을 가정하고, 다변량 분석이 가능한 VAR 모형을 활용한다. VAR은 한 번의 충격이 일정 기간 지속되는 시계열의 특성을 반영하며, 다변량의 관계가 시차를 두고 영향을 미치는 현상을 설명할 수 있다(Sims, 1980). 불확실성이 다른 시계열 변수에 주는 영향을 분석하는 연구도구로 VAR은 자주 활용되고 있다(Baker, et al., 2016 : Bachman, et al., 2013). 또한, 충격 반응 분석 (Impulse Response Analysis)을 통해, 하나의 변수가 변했을 때 다른 변수

3) “출산”이라는 키워드는 제외하였다. 그 이유는 “출산” 키워드만 들어간 기사는 출산에 대한 공포에 관한 기사뿐만 아니라 출산율 하락 자체에 관한 기사가 많이 포함되었기 때문이며, “육아”와 and 조건으로 포함시킬 필요가 적다고 판단되었기 때문이다.

4) 선별된 기사는 일부 내용이라도 출산/육아에 대해 부정적이면 족하다고 판단하였다. 즉, 기사의 의도와 취지가 반드시 “출산 육아는 힘들고 스트레스 받는다”는 논지일 필요는 없다. 예를 들어, 전체적으로는 어떤 인물의 성공 스토리에 대한 긍정적인 이야기이더라도, “과거에 육아에 너무 지치고 스트레스 받아서 모든 것을 포기하려고 하였다”라는 인터뷰 내용이 있으면 충분히 독자로 하여금 출산을 꺼리게 작용할 수 있기 때문이다.

5) 지표를 100을 기준으로 그보다 얼마나 오르고 내렸는지 직관적으로 판단할 수 있게 하는 것이 목적이다. 이를 위하여 지표 100의 기준은 어느 정도 의미 있는 기간인 10년을 기준 기간으로 정의하였다(이는 Baker의 방법론과 같다). 2004년부터 10년으로 정의한 이유는 2004년부터 인터넷 신문 매체들이 본격적으로 나타나기 시작하였기 때문이다.

에 어떤 영향을 미치는지 분석할 수 있다는 장점 때문에, 계량경제학의 시계열 분석 방법 중에서 가장 널리 사용되는 방법론 중의 하나이다. 먼저, 지표와 출산율의 단위근을 검정하고, VAR 모형에 적합한 시차를 도출한다. 그랜저 인과관계(Granger Causality) 분석을 통해 지표가 출산율에 영향을 미치는지 검정한다. 마지막으로 VAR 모형의 회귀 계수의 유의성을 검토하고, 충격 반응 함수까지 도출한다.

III. 분석 결과

분석은 2000년~2018년의 데이터를 대상으로 하였다. 그 이유는 2000년 이전에는 인터넷에 의한 뉴스 전파가 상대적으로 활발하지 못했기 때문이다. 2000년~2018년까지 육아에 대한 부정적인 기사를 기반으로 산출한 지표는 <표 3>과 <표 4>에 정리되어 있는데, <표 3>의 데이터는 언론사 전체⁶⁾를 포함한 결과이

〈표 3〉 육아에 대한 부정적인 기사 지표
〈Table 3〉 Negative News Index for Childcare

Year	A. childcare news (negative)	B. childcare news	C. social news	B/C	A/C	A/B	Index (A/C)
2000	81	522	27,942	1.9%	0.3%	15.5%	155
2001	123	782	40,800	1.9%	0.3%	15.7%	161
2002	195	1,034	87,031	1.2%	0.2%	18.9%	120
2003	240	1,192	172,090	0.7%	0.1%	20.1%	75
2004	438	2,128	2,006,430	0.1%	0.0%	20.6%	12
2005	639	3,073	2,421,524	0.1%	0.0%	20.8%	14
2006	985	4,640	1,872,480	0.2%	0.1%	21.2%	28
2007	1,170	5,487	1,866,061	0.3%	0.1%	21.3%	34
2008	1,196	6,570	1,221,159	0.5%	0.1%	18.2%	52
2009	1,734	9,050	1,897,025	0.5%	0.1%	19.2%	49
2010	2,423	13,715	1,291,294	1.1%	0.2%	17.7%	101
2011	2,519	15,044	798,564	1.9%	0.3%	16.7%	169
2012	3,962	22,817	975,603	2.3%	0.4%	17.4%	218
2013	5,380	36,282	889,533	4.1%	0.6%	14.8%	324
2014	8,885	59,195	1,202,065	4.9%	0.7%	15.0%	396
2015	11,145	73,854	1,404,737	5.3%	0.8%	15.1%	425
2016	12,414	74,196	1,352,771	5.5%	0.9%	16.7%	492
2017	15,797	80,907	1,577,389	5.1%	1.0%	19.5%	536
2018	19,321	85,825	1,776,461	4.8%	1.1%	22.5%	583

6) 일간지, 네이버에 콘텐츠를 제공하는 방송사, 인터넷 신문, 지역지 등 500여개의 매체 포함

고, <표 4>는 일간지⁷⁾만 포함하고 있다. 2004년 각종 인터넷 신문이 등장하면서 전체 기사의 양이 폭발적으로 증가하였다. 육아에 대한 기사는 2000년대 중반까지 줄어들다가, 최근 10년동안 지속적으로 늘어나는 추세를 볼 수 있다. 전체 사회면 대비 육아에 대한 부정적인 기사 지표도 최근에 크게 증가함을 알 수 있다.

표에는 앞서 설명한 지표 산출 과정에 따라 도출된 지표값들이 정리되어 있는데, 예를 들면, 2000년의

지표값 155는 2004년부터 2013년까지 평균 A/C값을 M이라고 하였을 때, M에 대한 상대적인 A/C값의 비중으로 정의된다. 즉, $A/C(t=2000년) \cdot \frac{100}{M}$ 으로 산출된 값이다.

이러한 결과는 일간지를 대상으로 한 분석(<표 4>)에서도 크게 다르지 않게 나타났다.

시계열 분석에 앞서 안정적인 (Stationary) 시계

<표 4> 육아에 대한 부정적인 기사 지표 (일간지)
<Table 4> Negative News Index for Childcare (Daily Newspapers)

Year	A. childcare news (negative)	B. childcare news	C. social news	B/C	A/C	A/B	Index (A/C)
2000	41	195	8,441	2.3%	0.5%	21.0%	116
2001	55	304	12,800	2.4%	0.4%	18.1%	102
2002	77	359	20,006	1.8%	0.4%	21.5%	92
2003	102	429	35,080	1.2%	0.3%	23.8%	69
2004	191	820	70,540	1.2%	0.3%	23.3%	65
2005	219	963	74,233	1.3%	0.3%	22.7%	70
2006	293	1,162	70,973	1.6%	0.4%	25.2%	98
2007	277	1,140	75,771	1.5%	0.4%	24.3%	87
2008	263	1,188	73,261	1.6%	0.4%	22.1%	86
2009	328	1,465	88,927	1.7%	0.4%	22.4%	88
2010	438	1,756	76,145	2.3%	0.6%	25.0%	137
2011	426	1,702	100,771	1.7%	0.4%	25.0%	101
2012	442	1,790	87,386	2.1%	0.5%	24.7%	121
2013	697	3,026	113,067	2.7%	0.6%	23.0%	147
2014	908	5,107	130,757	3.9%	0.7%	17.8%	166
2015	1,404	6,358	174,102	3.7%	0.8%	22.1%	192
2016	1,497	6,538	131,127	5.0%	1.1%	22.9%	272
2017	1,849	7,172	144,887	5.0%	1.3%	25.8%	304
2018	2,261	7,608	163,172	4.7%	1.4%	29.7%	331

7) 경향신문, 국민일보, 내일신문, 동아일보, 매일일보, 문화일보, 서울신문, 세계일보, 아시아투데이, 조선일보, 중앙일보, 한겨레, 한국일보

열 자료인지 여부를 판단하기 위하여 Augmented Dickey-Fuller 테스트(Dickey & Fuller, 1979)를 진행하였다. 귀무가설은 "시계열 데이터에 단위근(unit root)이 존재 한다"이고, 분석 결과, <표 3>의 Index(A/C)와 <표 4>의 Index(A/C)는 각각 p-value가 0.9984, 0.9950로서, 귀무가설을 기각하지 못한다. 즉, Index(A/C)는 단위근이 존재하는 안정적이지 않은 시계열 데이터이다. 안정적이지 않은 시계열 데이터를 활용하여 회귀분석을 진행할 경우에는 가성적(Spurious) 회귀가 될 수 있으므로, 이후 분석은 모두 로그차분(Log Difference) 통계량을 사용함으로써 이러한 문제를 해결하였다(Granger, et al., 2001). 임신을 하고 실제로 출산하기까지만 해도 1년 가까이의 기간이 소요되기 때문에, 1년 이상의 기간의 시차를 두고 상관관계 관계가 발생하는 것으로 보는 것이 타당하다. 이에 따라 시차를 달리하여(0~6년까지의 시차) 상관계수를 도출하였다.

<표 5>에서 볼 수 있듯이, 시차가 3년일 때 상관계수의 값이 가장 크다는 것을 알 수 있다. 특히 1아 출산의 경우 지표가 증가함에 따라 -0.704의 높은 상관관계를 나타내고 있는데, 이것은 육아에 대한 부정적인 언론이 높아지면, 3년 뒤 1아의 출산율이 낮아진다는 것을 의미한다.

가장 상관관계가 높은 시차 3년을 대상으로 지표와 출산율 간의 회귀분석 결과가 <표 6>에 정리되어 있는데, 지표의 증가에 따라 3년 뒤 출산율은 유의미하게 감소함을 나타내고 있다. 이러한 결과는 40대 이상의 여성에게는 유의하지 않으나, 20~30대 연령대에서는 유의하게 나타났다. 따라서, 40대 이상의 여성들과는 달리 20~30대 젊은 여성들은 육아에 대한 부정적인 기사에 영향을 받는 것으로 볼 수 있다.

또한, 지표와 출산아 간의 회귀 분석 결과가 <표 7>에 정리되어 있는데, 지표는 첫째아에 매우 유의한 영향을 미치는 것으로 나타났다. 이러한 결과는 <표 5>

<표 5> 출산율과 지표의 상관관계 (0~6년 지연시차 적용, 괄호는 p-value)
(Table 5) Correlation between Birthrate & Index (lag = 0~6, (p-value))

lag	dlog(birthrate)									
	total	20-24	25-29	30-34	35-39	40-44	45-49	1 child birth	2 children	3 & more
dlog(index(0))	0.205 (0.447)	0.429 (0.098)	0.079 (0.772)	0.026 (0.924)	0.291 (0.275)	0.442 (0.087)	-0.099 (0.716)	0.066 (0.809)	0.359 (0.172)	0.274 (0.304)
dlog(index(1))	0.396 (0.145)	0.558 (0.031)	0.403 (0.136)	0.180 (0.520)	0.316 (0.251)	0.465 (0.080)	-0.016 (0.956)	0.362 (0.185)	0.432 (0.108)	0.262 (0.346)
dlog(index(2))	-0.090 (0.759)	-0.004 (0.989)	-0.056 (0.849)	-0.245 (0.398)	-0.104 (0.722)	0.052 (0.859)	-0.037 (0.901)	-0.171 (0.558)	0.011 (0.969)	-0.033 (0.912)
dlog(index(3))	-0.615 (0.025)	-0.650 (0.016)	-0.570 (0.042)	-0.664 (0.013)	-0.622 (0.023)	-0.436 (0.137)	-0.141 (0.646)	-0.704 (0.007)	-0.463 (0.111)	-0.432 (0.140)
dlog(index(4))	0.212 (0.509)	-0.041 (0.899)	0.235 (0.872)	0.052 (0.872)	0.195 (0.544)	0.229 (0.474)	0.180 (0.576)	0.204 (0.525)	0.137 (0.670)	0.127 (0.693)
dlog(index(5))	0.094 (0.783)	0.102 (0.766)	0.011 (0.840)	-0.069 (0.840)	0.137 (0.689)	0.019 (0.956)	0.084 (0.806)	0.097 (0.778)	-0.044 (0.899)	0.081 (0.813)
dlog(index(6))	-0.251 (0.484)	-0.171 (0.636)	-0.173 (0.227)	-0.420 (0.227)	-0.436 (0.208)	-0.503 (0.138)	-0.077 (0.932)	-0.068 (0.851)	-0.439 (0.205)	-0.672 (0.033)

〈표 6〉 연령별 출산율과 지표(3)의 단순 회귀 분석 결과
 〈Table 6〉 Regression between Birthrate by age & Index(3)

independent var: index	dependent var: dlog(birthrate), (p-value)						
	Total	20~24	25~29	30~34	35~39	40~44	45~49
dlog(index(3))	-0.053 (0.025)	-0.067 (0.016)	-0.053 (0.042)	-0.062 (0.013)	-0.055 (0.023)	-0.040 (0.1368)	-0.591 (0.646)

〈표 7〉 출산순위별 출산아 수와 지표(-3)의 단순 회귀 분석
 〈Table 7〉 Regression between # of Children at Birth & Index(3)

independent var: index	dependent var: dlog(# of children at birth), (p-value)		
	1	2	3 & more
dlog(index(3))	-0.066 (0.007)	-0.040 (0.111)	0.049 (0.141)

에서 살펴본 분석 결과와 유사한데, 즉, 육아에 대한 부정적인 기사가 출산을 하지 않은 부모에게는 영향을 미치지 않지만, 일단 최소 한명의 아이를 가지고 있는 부모에게는 유의한 영향을 미치지 않는 것으로 볼 수 있다.

다음으로는 VAR 모형을 분석한다. VAR 모형은 다음의 식 (1)과 같이 정의할 수 있다.

$$X_t = A_0 + A_1 \cdot X_{t-1} + A_2 \cdot X_{t-2} + \dots + A_p \cdot X_{t-p} + \epsilon_t \quad \text{식 (1)}$$

여기에서 변수 $\begin{pmatrix} \delta \log(\text{출산율})_t \\ \delta \log(\text{지표})_t \end{pmatrix}$ 는 이며,

$A_i (i=0 \sim p)$ 는 2×2 행렬이다.

〈표 8〉 VAR 시차 선택
 〈Table 8〉 VAR Lag Selection

Lag	LogL	LR	FPE	AIC	SC	HQ
0	17.946	NA*	0.00024	-2.658	-2.577	-2.688
1	19.998	3.079	0.00034	-2.333	-2.091	-2.423
2	24.728	5.518	0.00033	-2.455	-2.051	-2.604
3	33.839	7.593	0.00018*	-3.307*	-2.741*	-3.516*
4	36.620	1.390	0.00038	-3.103	-2.376	-3.373

* : selected lag based on statistical criteria

LR: sequential modified LR test statistics (at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hanna-Quinn information criterion

8) 앞서 상관계수 분석 결과에서 적절한 시차는 3이라는 것을 알 수 있으며, 이에 따라 시차를 4까지 고려하였다. VAR에서 시차를 6까지 고려하는 경우에는 구해야 하는 회귀계수가 상수항 포함할 경우 14개까지 늘어나기 때문에 모형의 자유도가 낮아지는 문제가 있다.

$\epsilon_t = \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}$ 으로 백색잡음 (white noise)이다. 여기서 p 는 VAR 모형의 시차로서, VAR 분석을 위해서 적합한 시차를 먼저 결정해야 한다. 시차를 0~4까지 변경 하면서⁸⁾ 검증한 결과는 다음과 같다(〈표 8〉).

〈표 8〉에서 볼 수 있는 바와 같이, FPE, AIC, SC, HQ 등의 네 가지 기준에서 공통적으로 시차=3이 적합하다고 판단되어, VAR(3) 모형을 선택하였다. 본격적인 VAR 분석에 앞서 그랜저 인과관계 분석을 수행하였다(Granger, 1969). 그랜저 인과관계 분석을 수행한 이유는 지표의 증감이 출산율에 영향을 미치지 않지만, 반대로 출산율의 증감이 지표에 영향을 미치지 않는 것이 일반적인 직관에 맞기 때문이다. 즉, 지표의 증감과 출산율의 관계가 단순 상관관계를 넘어서 인과관계까지 존재하는지 여부를 실증하는 것이 필요하다고 판단하였다. 지표가 출산율에 영향을 미치지 않는다는 (그랜저 인과관계가 없다는) 귀무가설을 검정하기 위하여 식(2)와 식(3)을 비교하여 추가적인 항들이 회귀 모형의 설명력을 증대시키는지 테스트한다.

$$\delta \log(\text{출산율})_t \quad \text{식 (2)}$$

$$= a_0 + a_1 \delta \log(\text{출산율}_{t-1}) + a_2 \delta \log(\text{출산율}_{t-2}) + a_3 \delta \log(\text{출산율}_{t-3})$$

$$\delta \log(\text{출산율})_t \quad \text{식 (3)}$$

$$= a_0 + a_1 \delta \log(\text{출산율}_{t-1}) + a_2 \delta \log(\text{출산율}_{t-2}) + a_3 \delta \log(\text{출산율}_{t-3}) + b_1 \delta \log(\text{지표}_{t-1}) + b_2 \delta \log(\text{지표}_{t-2}) + b_3 \delta \log(\text{지표}_{t-3}) + \epsilon_t$$

〈표 9〉에서 볼 수 있는 바와 같이, 지표가 출산율에

〈표 9〉 그랜저 인과관계 테스트 결과
〈Table 9〉 Granger Causality Test Results

Null Hypothesis	F stat.	Prob.
dlog(index) does not affect dlog(birthrate)	5.126	0.043
dlog(birthrate) does not affect dlog(index)	2.064	0.207

영향을 미치지 않는다는 귀무가설은 기각되고, 출산율이 지표에 영향을 미치지 않는다는 귀무가설은 기각되지 않았다. 따라서, 지표는 출산율에 영향을 미치지 않지만, 역으로 출산율은 지표에 영향을 미치지 않음을 알 수 있다. 이것은 언론에서 언급되는 출산 육아에 대한 부정적인 기사가 실제 출산율에는 영향을 미치지 않지만, 출산율이 증감하는 사실이 언론의 기사에 영향을 미치지 않는 못함을 의미한다.

〈표 10〉은 합계 출산율과 지표의 VAR(3) 분석 결과이다. 인과관계 분석 결과, 지표 변화가 출산율에 영향을 미치고 역은 성립하지 않으므로, 종속변수가 지표, 독립변수가 출산율인 경우를 검토한다. 그 결과 3년 전에 지표의 1단위 증가당 출산율의 변화율은 -0.051임을 알 수 있다. 즉, 출산에 대한 부정적인 언급이 늘어지면 3년 뒤 출산율은 떨어지며 이는 상관관계가 아닌 그랜저 인과관계가 있다는 의미이다.

〈표 10〉 합계출산율, 지표의 VAR 분석 결과
〈Table 10〉 VAR Analysis Results of Total Birthrate & Index

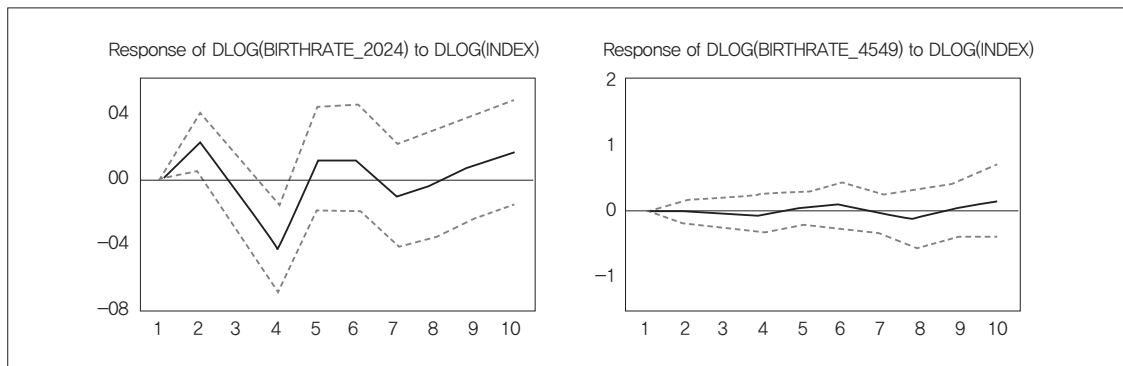
	coefficient, standard error, T statistics	
	dlog(birthrate)	dlog(index)
dlog(birthrate(1))	-0.502 -0.261 [-1.920]	6.551 -4.672 [1.402]
dlog(birthrate(2))	-0.327 -0.183 [-1.785]	5.937 -3.271 [1.815]
dlog(birthrate(3))	0.028 -0.194 [0.145]	7.661 -3.469 [2.208]
dlog(index(1))	0.045 -0.018 [2.514]	-0.133 -0.320 [-0.414]
dlog(index(2))	0.020 -0.020 [1.010]	-0.375 -0.351 [-1.070]
dlog(index(3))	-0.051 -0.017 [-3.090]	-0.248 -0.297 [-0.835]

연령대별 출산율이 지표에 얼마나 반응하는지 분석한 결과는 다음의 <표 11>에 정리되어 있다. 20대 여성과 30대 여성은 지표(3)에 음의 방향으로 유의미한

반응을 보이며 특히 20대가 더 유의미한 반응을 보이고 있다. 반면 40대 이후는 지표의 변화에 유의미한 반응을 보이지 않고 있다. 즉, 출산에 대한 부정적인

<표 11> 연령대별 출산율, 지표의 VAR 분석 결과
(Table 11) VAR Analysis Results of Birthrate by Age & Index

	coefficient, standard error, T statistics					
	20~24		30~34		40~44	
	dlog(birthrate)	dlog(index)	dlog(birthrate)	dlog(index)	dlog(birthrate)	dlog(index)
dlog(birthrate(1))	0.212 0.210 [1.008]	8.385 3.671 [2.284]	-0.304 0.315 [-0.964]	0.980 4.196 [0.234]	0.148 0.349 [0.425]	7.287 3.568 [2.042]
dlog(birthrate(2))	-0.181 0.157 [-1.152]	1.712 2.751 [0.622]	-0.093 0.271 [-0.342]	3.399 3.605 [0.943]	-0.285 0.327 [-0.871]	1.768 3.349 [0.528]
dlog(birthrate(3))	0.368 0.149 [2.468]	7.753 2.611 [2.969]	0.100 0.263 [0.380]	4.514 3.502 [1.28909]	0.346 0.305 [1.134]	7.441 3.124 [2.38207]
dlog(index(1))	0.046 0.017 [2.647]	-0.352 0.306 [-1.151]	0.020 0.027 [0.729]	0.094 0.359 [0.262]	0.041 0.032 [1.290]	-0.160 0.326 [-0.493]
dlog(index(2))	-0.013 0.020 [-0.639]	-0.747 0.350 [-2.132]	-0.005 0.027 [-0.181]	-0.042 0.363 [-0.116]	-0.003 0.035 [-0.091]	-0.598 0.356 [-1.680]
dlog(index(3))	-0.073 0.017 [-4.414]	-0.326 0.290 [-1.124]	-0.066 0.027 [-2.487]	-0.057 0.354 [-0.161]	-0.045 0.031 [-1.448]	-0.310 0.317 [-0.978]



<그림 3> 20대 초반, 40대 후반의 지표에 대한 충격 반응
<Fig. 3> Impulse Response of early 20's and late 40's toward Index

언급이 늘어나면 20대의 출산율에는 유의미한 영향을 미치지 않지만, 나이가 많아질수록 이러한 영향은 적어지거나 없어진다는 것을 알 수 있다.

지표의 변화에 따른 연령별 변화는 충격반응에서 더욱 잘 확인할 수 있다. <그림 3>에서 볼 수 있듯이 20~24세 여성의 출산율의 경우 지표의 충격에 대하여 3기 뒤 ($t=4$)에 유의미하게 떨어지는 반면, 45~49세 여성의 경우는 지표에 거의 반응하지 않는다. 이는 <표 11>의 결과와 일맥상통하며, 20대의 경우 출산에 대한 부정적인 언급이 늘어날 경우 3년 후에 출산율이 떨어지지만, 나이가 많아질수록 그러한 영향은 없어지는 것을 볼 수 있다.

IV. 결론

본 연구에서는 2000년~2018년까지 육아에 관한 부정적인 기사가 전체 사회 기사 대비 차지하는 비중을 기반으로 출산에 대한 부정적 언급에 대한 지표를 산출하였다.

본 연구의 주요한 분석 결과를 정리하면 다음과 같다.

첫째, 육아에 관한 부정적인 기사에 대한 지표의 변화는 출산율에 영향을 미치지 않지만, 반대로 출산율의 변화는 지표에 변화에 영향을 미치지 않는 것으로 나타났다. 즉, 부정적인 기사의 증가는 출산율의 감소에 상관관계가 아닌 인과관계를 가지고 있다는 것이다.

둘째, 지표가 증가하면 3년 뒤의 출산율이 감소하는 것으로 나타났다. 나이가 어릴수록 지표에 대한 반응은 크게 나타났는데, 20대 초반 여성의 출산율은 지표 증가 후 3년 뒤 유의미한 감소 반응을 나타내는 반면, 45세 이상 여성의 출산율은 지표에 거의 반응하지 않는 것으로 나타났다. 전체적으로 연령대가 어릴수록 지표에 유의미한 영향을 받는 것을 의미한다.

셋째, 지표는 1아 출산율에 유의미한 영향을 미치는 것으로 나타났다. 출산율을 높이기 위하여 가장 중요한 1아 출산율에 영향을 미친다는 것은 의미가 큰 결

과이다. 또한 이미 출산과 육아를 경험해본 사람들에게는 육아에 대한 부정적인 기사가 큰 영향을 미치지 못한다는 것을 의미한다.

그동안 뉴스 빅데이터의 활용은 형태소 분석 및 키워드 빈도 모니터링 수준에서 크게 벗어나지 않았으며, 사회 현상을 예측하는데 활발하게 활용되지 못하였다. 그 이유는 뉴스 빅데이터는 비정형적이며 정제하여 분석하기가 쉽지 않고, 의미를 도출하기에는 노이즈 정보가 훨씬 많기 때문이다. 본 연구는 뉴스 빅데이터를 이용하여 단순한 키워드 빈도 모니터링에 그치지 않고, 사회 현상을 예측할 수 있음을 보인 것에 의의가 있다고 판단된다. 또한, 빅데이터 기반의 지표를 활용하는 것은 실시간성이 있는데, 본 연구에서 제시한 지표는 출산율이 떨어질 수 있는 징후를 미리(3년 선행) 파악하는데 기여할 수 있을 것이다.

일반적으로 시계열의 VAR 분석 결과는 유의미한 결과가 나오는 경우가 많지 않은데, 그 이유는 첫째, 시계열 데이터는 시간이 지남에 따라 설명변수가 달라지거나 메커니즘이 달라지는 경우가 많으며, 둘째, 계절적 요인, 시대적 요인 등 설명변수에 반영할 수 없는 정보가 훨씬 많기 때문이다. 이러한 관점에서 언론의 육아에 대한 부정적인 언급이 심리적인 면으로 작용하여 출산율 저하로 연결되는 현상은 무시할 수 없다고 판단된다. 향후 출산율에 영향을 미치는 키워드를 추가적인 검증 절차를 거쳐서 뽑아낸다면, 더욱 유의미한 결과를 얻을 수 있을 것이라고 생각한다. 한편, 본 연구에서는 2000년 데이터부터 분석을 하였지만, 향후 시간이 지나면서 빅데이터가 좀 더 축적된다면 다른 변수들도 함께 고려한 분석이 가능할 것으로 생각한다.

또한 본 연구에서는 연단위의 시계열 데이터를 활용하였으나 향후 분기 단위 데이터를 활용하면 보다 자유도 높은 모형 기반으로 정교한 분석이 가능할 것으로 판단된다. 이 경우 트렌드나 특정 성분 처리 등 여러 가지 시계열 데이터 분석 기법이 활용되어야 할 것으로 생각한다.

■ References

- Angeles, L. (2010). Demographic Transitions: Analyzing the Effects of Mortality on Fertility. *Journal of Population Economics*, 23(1), 99-120.
- Arbatli, E. C., Davis, S. J., Ito, A., Miake, N. & Saito, I. (2017). *Policy Uncertainty in Japan* (No. w23411). National Bureau of Economic Research.
- Bachmann, R., Elstner, S. & Sims, E. R. (2013). Uncertainty and Economic Activity: Evidence from Business Survey Data. *American Economic Journal: Macroeconomics*, 5(2), 217-249.
- Baker, S. R., Bloom, N. & Davis, S. J. (2016). Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636.
- Choi, J., Jang, J. & Kim, B. (2014). The Determinants for the Preschool's Comprehensive Information Systems: User Resistance Perspective. *Informatization Policy*, 21(1), 77-98.
- {최재원·장지화·김범수 (2014). 유아교육 정보시스템 활성화를 위한 결정요소-사용자 저항성의 관점으로. <정보화정책>, 21권 1호, 77-98.}
- Del Bono, E., Weber, A. & Winter-Ebmer, R. (2012). Clash of Career and Family: Fertility Decisions after Job Displacement. *Journal of the European Economic Association*, 10(4), 659-683.
- Dickey, D. A. & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American statistical association*, 74(366a), 427-431.
- Granger, C. W. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica: Journal of the Econometric Society*, 37(3), 424-438.
- Granger, C. W., Newbold, P. & Econom, J. (2001). Spurious Regressions in Econometrics. *A Companion to Theoretical Econometrics*, Blackwell, Oxford, 557-561.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton university press.
- Hong, J. (2013). The Analysis of the Effectiveness of Governmental Child-care Subsidies. *Korea Journal of Population Studies*, 36(4), 95-118.
- {홍정립 (2013). 보육비 지원 정책의 효과성 분석. <한국 인구학>, 36권 4호, 95-118.}
- Hong, M. (2009). Information Technology for Aged Welfare-Focus on Gratification of the Users-. *Informatization Policy*, 16(1), 75-86.
- {홍명신 (2009). 노인정보복지 실천 연구-인터넷 충족을 중심으로-. <정보화정책>, 16권 1호, 75-86.}
- Hong, S. C., Kim, Y. I., Lim, J. Y. & Yeo, M. Y. (2016). Pro-natalist Cash Grants and Fertility: A Panel Analysis. *The Korean Economic Review*, 32(2), 331-354.
- Jurado, K., Ludvigson, S. C. & Ng, S. (2015). Measuring Uncertainty. *American Economic Review*, 105(3), 1177-1216.
- Kim, M., Lee, W., Lee, H. & Suh, B. (2018). Forecasting the Future Korean Society: A Big Data Analysis on 'Future Society'-related Keywords in News Articles and Academic Papers. *Informatization Policy*, 21(1), 37-64.
- {김문조·이왕원·이혜수·서병조 (2018). 빅데이터를 통해 본 한국사회의 미래: 언론사 뉴스기사와 사회과학 학술논문의 '미래사회' 관련 키워드 분석. <정보화정책>, 25권 4호, 37-64.}
- Kim, S., Choi, B., Chung, K., Lee, S., Park, D., Park, I. & Chang, J. (2002). *Social and Economic Impact and Long & Short-term Policy for Low Fertility*. Korea Institute for Health and Social Affairs.
- {김승권·최병호·정경희·이삼식·박덕규·박인화·장지연 (2002). <저출산의 사회경제적 영향과 장단기 정책방안(정책보고서2002-46)>. 보건복지부, 한국보건사회연구원.}
- Kim, S., Kim, M., Lee, Y. Sunwoo, D., Cho, S. & Kim, O. (2003). *Research on Population Policy and Inter-governmental Implementation Scheme Development Against Low Fertility*. Korea Institute for Health and Social Affairs.
- {김승권·김민자·이연주·박정환·선우덕·조성현·강옥희 (2003). <저출산 대비 인구정책 개발 및 범정부 추진 체계 수립 연구>. 보건복지부, 한국보건사회연구원.}

- Künzler, J. (2002). Paths Towards a Modernization of Gender Relations, Policies, and Family Building. *Family Life and Family Policies in Europe. Problems and Issues in Comparative Perspective*, 2, 252-298.
- Langridge, A. T., Nassar, N., Li, J., Jacoby, P. & Stanley, F. J. (2012). The Impact of Monetary Incentives on General Fertility Rates in Western Australia. *J Epidemiol Community Health*, 66(4), 296-301.
- Lee, B. (2003). Gender Discrimination and Family Welfare Policy of Low Fertility. *National Assembly Library Monthly Magazine*, 40(10), 63-79.
- {이병렬 (2003). 저출산의 성차별적 원인과 가족복지정책. <국회도서관보>. 40권 10호, 63-79.}
- Lee, M. & Myung, S. (2015). A Study on the Effectiveness of the Childbirth Promotion Policy. *Korean Journal of Local Government & Administration Studies*, 29(1), 331-350.
- {이미옥·명성준 (2015). 출산장려정책의 효과성에 대한 연구. <한국자치행정학보>, 29권 1호, 331-350.}
- Lee, S. (2006). A Study on Impact of the Change in Values on Marriage and Fertility Behaviors. *Health and Social Welfare Review*, 26(2), 96-140.
- {이삼식 (2006). 가치관의 변화가 결혼 및 출산 행태에 미치는 영향. <보건사회연구>. 26권 2호, 96-140.}
- Lovenheim, M. F. & Mumford, K. J. (2013). Do Family Wealth Shocks Affect Fertility Choices? Evidence from the Housing Market. *Review of Economics and Statistics*, 95(2), 464-475.
- Ministry of Women (2002). Statistical Yearbook on Women, Ministry of Women.
- {여성부 (2002). <여성통계연보>. 여성부.}
- Nimark, K. P. (2014). Man-bites-dog Business Cycles. *American Economic Review*, 104(8), 2320-67.
- Shin, Y. (2008). An Analysis of the Effect of Burden of Childcare and Education on Childbirth. *Health and Social Welfare Review*, 28(2), 103-134
- {신윤정 (2008). 보육 양육비 부담이 출산 의향에 미치는 영향 분석. <보건사회연구>, 28권 2호, 103-134.}
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica: journal of the Econometric Society*, 48(1), 1-48.
- Statistics Korea (2018). *2018 Birth Statistics (Finalized)*. Statistics Korea.
- {통계청 보도자료 (2018). <2018년 출생 통계(확정)>. 통계청.}