

임금 및 소득이 출산 간격에 미치는 효과의 분석

민 희 철*

요 약 문

본 연구는 여성 임금과 배우자 소득수준이 결혼 후 첫 번째 및 두 번째 출산간격에 미치는 효과를 1998년 이후 결혼한 여성에 대해 분석하였다. 출산 결정을 모형화하기 위하여 듀레이션 모형(duration model)을 사용하였으며, 임금 및 소득에 대한 정보가 부족한 출산력 자료를 보충하기 위하여 노동패널자료로부터 추정된 임금 방정식을 이용하였다. 분석 결과 여성 임금의 증가가 첫 번째 출산을 지연시키는 효과를 발견하였는데, 추정된 모형에 의하면 여성 임금의 20% 증가는 표본 중 50%가 출산을 완료하는 시점을 기준으로, 첫 출산의 경우 2개월, 두 번째 출산의 경우 1개월 정도 지연시킬 것으로 예측되었다. 반면 배우자 소득이 출산을 앞당기는 효과는 첫 번째 출산의 경우 거의 나타나지 않았고, 두 번째 출산의 경우에만 통계적으로 유의한 것으로 나타났다.

핵심주제어: 출산 간격, 출산율, 듀레이션 모형

JEL 번호: J10, J13, C41

I. 도입

지난 2005년 우리나라의 합계출산율은 1.08이라는 유례없이 낮은 수준을 기록하면서 저출산의 문제가 우리사회의 매우 심각한 문제로 본격적으로 주목받게 되었다. 1960년 6.0의 높은 수준이던 출산율은 지속적으로 감소하여 1983년에 대체출산율(replacement rate)인 2.1 이하로 떨어졌으며, 2000년 이후에는 1.3 이하의 매우 낮은 수준에 머물러 있다. 이로 인하여 매년 새로 태어나는 출생아의 숫자도 1960년 100만 명을

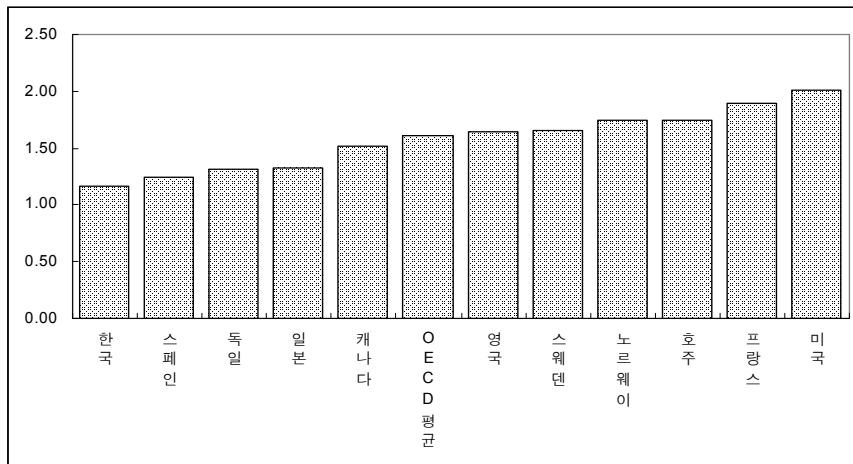
* 한국조세연구원 전문연구위원, e-mail: min@kipf.re.kr

논문투고일: 2007년 8월 2일, 심사완료일: 2008년 2월 11일

넘는 수준이었던 것이 2005년에는 44만명으로 절반 이상 줄어들었다. 우리나라의 출산율은 다른 국가와 비교하여도 매우 낮은 수준인데, 2002년 주요한 OECD 국가의 합계출산율을 그래프로 나타낸 [그림 1]을 통해서 확인할 수 있다. 이에 따르면 2002년 당시 우리나라의 합계출산율 1.17은 OECD 평균인 1.60은 물론, 스페인의 1.25, 독일 1.31, 일본 1.32보다 훨씬 낮아서 OECD 국가 중에서도 최하위에 속하는 것을 알 수 있다. 즉 최근의 저출산은 과거 우리나라는 물론 비슷한 시기 다른 국가와 비교하여도 유례를 찾을 수 없을 정도의 심각한 수준에 이르렀다.

정부에서도 저출산 문제의 심각성을 인식하고 출산 및 양육에 유리한 환경을 조성하기 위하여 ‘저출산고령사회기본계획’을 작성하고 2006~2010년에 걸쳐 총 32.1조원을 투자할 계획을 수립하는 등 대책 마련에 고심하고 있다. 그런데 정부가 정책을 통하여 출산에 의도된 영향을 미치기 위해서는 우선 출산이 무엇에 의하여 어떻게 결정되는지에 대한 명확한 이해가 선행되어야 한다. 물론 자녀를 가지는 결정에 영향을 미치는 원인이 단순히 경제적 요인에만 국한되지는 않을 것이다. 그러나 아동수당의 지급, 보육비용에 대한 세금공제 및 보조금 지원 등 주요한 출산장려정책이 기본적으로는 출산에 대한 경제적 인센티브를 제공한다는 점에서 출산의 경제적 요인을 파악하는 것이 매우 중요하다.

[그림 1] 주요 OECD 국가의 합계출산율 (2002년)



자료: OECD(2006)

이러한 문제의식을 가지고 본 연구에서는 여성의 출산에 영향을 미치는 두 가지 중요한 요인으로 여성의 임금과 배우자의 소득에 주목하였다. 특히 두 변수는 출산에 대한 경제학적 모형의 기본 구성요소라는 점에서 우선적으로 검토할 필요가 있다. 여성의 임금은 출산에 따르는 여성의 기회비용을 나타내는 변수이고, 배우자의 소득은 여성의 노동시장 참여와는 독립적인 소득으로서 출산의 소득효과를 측정할 수 있게 하는 변수이다.

기존 국내의 출산관련 실증연구는 대체로 인구사회학적 요인의 분석에 주목한 반면, 경제적 요인의 효과에 대해 충분히 검토하지 못하였다. 그 이유 중 하나는 출산관련 데이터에 노동시장관련 정보가 충분히 제공되지 않았던 데에 있다. 예를 들어 '2003년 전국 출산력 및 가족보건·복지실태 조사'는 출산 및 가족에 대한 매우 자세한 내용을 담고 있으나 임금, 근로형태 등 노동시장관련 내용은 상대적으로 부족하다. 이러한 문제를 해결하기 위하여 본 연구는 노동패널자료를 이용하여 임금방정식을 추정하고 그 결과를 출산력자료와 결합하여 사용함으로써 여성의 임금 및 배우자의 소득이 출산에 미치는 효과를 분석할 수 있었다. 분석 결과에 따르면 여성 임금의 증가는 첫 번째 출산과 두 번째 출산을 모두 유의미하게 지연시키는 효과를 갖는 것으로 나타났는데, 여성 임금의 20% 증가는 첫 출산을 2개월 정도, 두 번째 출산은 1개월 정도 지연시킬 것으로 추정되었다. 반면 배우자 소득이 출산을 앞당기는 효과는 두 번째 출산에 대해서만 유의성이 있는 것으로 나타났다. 이러한 결과는 출산에 관한 신가계경제학(New Home Economics)의 예측이 우리나라의 사례에서도 대체로 유효함을 의미한다.

논문의 구성은 다음과 같다. 우선 II절에서 관련된 선행연구를 검토하였으며, III절에서는 실증모형을 설명하였다. IV절에서는 실증연구에서 사용한 자료의 구성 및 기술적(descriptive) 통계량을 제시하였고, V절은 실증모형을 추정한 결과와 그에 대한 해석을 제공하였다. 마지막으로 VI절에서는 주요한 결론을 요약하고 앞으로의 연구 방향에 대해 설명하였다.

II. 선행 연구

결혼 및 출산에 영향을 주는 경제적 요인에 대한 본격적인 분석이 시작

된 것은 Becker(1960) 이후 신가계경제학(New Home Economics)으로 불리는 연구 결과로부터이다. 이에 따르면 자녀는 부모에게 효용을 주는 일종의 재화로 볼 수 있으며, 소비자 이론의 분석방법을 이용하여 출산 결정을 부모의 합리적인 선택으로 모형화한다. 따라서 자녀를 출산하고 양육하는데 소요되는 비용, 가구의 소득 등이 자녀수를 결정하는 주요한 변수가 된다. 특히 출산 및 양육 관련 비용에는 식품비, 의료비, 교육비 등 직접적으로 자녀를 위하여 지출하는 비용 이외에도 양육을 담당하는 여성이 근로하지 못함으로써 발생하는 기회비용이 포함된다는 점에 주목하여, 여성 임금의 상승이 출산에 부정적인 효과를 미친다는 예측을 제공하였다. 반면 자녀가 일반적인 재화와 마찬가지로 소득이 높을수록 더 많이 수요하는 성격을 지니는 정상재(normal goods)라면, 여성의 근로와 독립적인 소득의 증가, 예를 들어 배우자 소득의 증가는 출산 활동의 증가로 나타날 것으로 예측되었다.

Heckman and Walker(1990)는 여성 임금과 배우자 소득이 출산의 지연에 미친 효과를 듀레이션 모형을 이용하여 분석한 선구적인 연구이다. 이들은 스웨덴의 출산력 자료에 대한 분석을 통하여 임금 및 소득의 기본적인 경제 변수를 이용한 간단한 출산 모형이 인구사회학적 변수만을 고려한 모형에 비하여 출산율에 대한 설명력이 더 높다는 점을 보여주었다. 또한 여성 임금이 출산율을 낮추는 효과와 배우자 소득이 출산율을 높이는 효과를 확인하고, 절대적인 크기에서 여성 임금의 효과가 배우자 소득의 효과보다 크다는 것을 발견하였다. Merrigan and St.-Pierre (1998)은 유사한 모형을 이용하여 1950년부터 1990년까지 캐나다의 출산 활동을 분석하고 비슷한 결론을 도출하였다. 앞의 두 연구가 신가계경제학의 예측에 대체로 부합하는 결론을 보여준 실증연구인 반면, 반대의 근거를 보여준 연구도 존재한다. 예를 들어 Tasiran(1995)은 Heckman and Walker(1990)가 사용한 여성 임금 및 배우자 소득이 연도별 남성 소득과 여성 임금의 평균값을 사용한 점을 문제로 지적하고, 개인 수준의 자료를 보충하여 모형을 다시 추정할 경우 여성 임금이 출산율에 미치는 음의 효과가 유의미하지 않다고 주장하였다.

출산 시점 결정에 관한 국내 연구는 주로 인구사회학적 요인의 효과에 집중하였다는 특징이 있다. 은기수(2001)는 1997년 전국 출산력 조사 자료를 이용하여, 결혼 연령이 증가할수록 첫 출산간격이 짧아지지만, 최근 결혼코호트로 올수록 첫 출산간격이 늘어나고 있음을 보여주었다. 이

삼식 외(2005)는 2005년 전국 출산력 조사를 활용하여 첫 출산간격은 물론 두 번째 및 세 번째 출산간격의 결정요인을 각각 분석하였고, 미혼 여성의 결혼 시점에 대한 연구도 포함하고 있다. 출산시점을 결정하는 요인으로 교육수준, 초혼연령, 임신중절경험, 연령 등 다양한 인구사회학적 변수들의 효과가 분석되었다. 그러나 앞의 두 연구 모두 인구사회학적인 요인 이외에 여성 임금 및 배우자 소득 등 경제적 변수의 역할이 분석되지 못하였다. 그 주요한 이유는 연구의 관심사가 경제적 요인보다는 인구학적 요인에 있기 때문이겠으나, 한편으로는 두 연구가 사용한 자료가 갖는 제약 때문이기도 하다. 전국 출산력 자료는 조사시점에서 피응답자에 대한 회고적 서베이(retrospective survey)를 통해 정보를 수집하기 때문에 조사시점의 소득 및 임금 정보만을 제공한다. 즉 출산 결정에 영향을 미쳤던 출산 당시의 경제적 변수를 관찰하는 것이 불가능하므로 그 효과를 분석하는 것이 불가능하다. 본 연구는 노동패널자료를 활용하여 추정된 여성 임금 및 남성 소득 방정식을 출산력 자료와 결합하여 경제적 요인이 출산에 미치는 효과를 실증적으로 분석하였다. 이러한 분석은 기존의 출산 결정에 관한 인구사회학적 연구가 충분히 다루지 못한 부분으로서 본 연구 나름의 의의가 있다고 본다.

한편 류덕현(2007)은 집계변수를 사용한 거시적 분석을 통해 신가계경제학의 가설을 검증하였는데, 분석 결과 여성의 노동시장임금의 상승이 출산율에 미치는 효과는 젊은 연령에서 뚜렷하였고, 남성 소득의 효과는 높은 연령층에서 더 민감한 것으로 나타났다. 본 연구는 개인 수준의 미시 자료를 이용한 분석이라는 점에서 류덕현(2007)과 서로 보완적인 연구가 될 수 있을 것이다.

III. 실증 모형

만약 표본의 모든 여성에 대해 모든 출산을 관찰할 수 있다면 출산간격을 피설명변수로 하여 통상적인 회귀분석을 쉽게 수행할 수 있다. 그러나 관찰이 종료되는 시점까지도 출산이 완료되지 않는 경우 출산간격이 절단(censoring)되는 문제가 발생하며, 듀레이션 분석에서와 같이 출산간격의 분포를 명시적으로 고려하여야만 모형의 추정이 가능하다.

출산간격의 분포는 위험함수(hazard function)와 생존함수(survival function)로 정의하는 것이 일반적이다. 본 연구에서도 Heckman and Walker (1990), Tasiran(1995), Klasen and Launov(2006) 등 선행연구와 마찬가지로 와이불(Weibull) 분포의 위험함수 h 를 다음과 같이 정의하였다.¹⁾

$$h_j(t_j|Z_j, \omega) = \exp(\delta_j + \gamma_j \ln(t_j) + Z_j \beta_j + \omega) \quad (1)$$

j 는 출산순서를 나타내는 인덱스이며 t_j 는 j 번째 출산간격을 의미한다. 출산간격을 결정하는 요인은 Z 와 ω 로 나타내었는데, Z 는 연구자가 관찰할 수 있는 특성을 나타내며, ω 는 관찰되지 않는 이질성(unobserved heterogeneity)을 의미한다. 관찰되지 않는 이질성이란 예를 들어 여성의 출산능력(fecundity)을 나타내는 것으로 개인에 따라 출산이 빨리 이루어지거나 그렇지 않은 차이를 모형에서 반영하기 위한 구성요소이다. γ 는 듀레이션 의존성(duration dependence)을 나타내는 파라미터로, 시간이 지남에 따라서 위험함수가 증가하거나 또는 감소하는지를 결정한다.

생존함수 $S(t)$ 는 t 시점까지 출산이 이루어지지 않을 확률, 즉 출산간격이 t 보다 길 확률로 정의되며 다음과 같이 나타낼 수 있는데,

$$\begin{aligned} S_j(t_j|Z_j, \omega) &= p_j + (1-p_j) \exp\left(-\int_0^{t_j} h_j(s|Z_j, \omega) ds\right) \quad (2) \\ &= p_j + (1-p_j) \exp\left(-\exp(\delta_j + Z_j \beta_j + \omega) \frac{1}{\gamma_j + 1} t_j^{\gamma_j + 1}\right) \end{aligned}$$

이 때, 적절한 조건 하에서²⁾ 생존함수 S_j 는 t 가 증가함에 따라 p_j 에 수렴한다. 즉 p_j 는 시간이 충분히 지난 이후에도 j 번째 출산이 이루어지지 않을 확률을 의미하므로, 곧 $j-1$ 번째 출산 이후 출산을 중단할 확률로 해석할 수 있다.³⁾ 다만 출산 중단이 자발적인 선택 때문인지 아니면

1) 편의상 개별 관찰값을 나타내는 인덱스 j 는 생략하였다.

2) γ 가 -1보다 작은 값을 가지지 않는다는 조건

3) p 가 0과 1사이의 값을 가져야 한다는 제약이 최우추정을 용이하지 않게 만들므로, 실제 추정에서는 $p_j = 1/(1 + \exp(-\mu_j))$ 로 변환(reparametrize)하여서 제약 없이 μ 를 추정하였다.

불임 등의 외부적 요인으로 인한 것인지 모형에서 구분하고 있지 않으므로 p_j 의 해석에는 주의가 필요하다.

식 (2)의 생존함수는 관찰할 수 없는 여성의 특성 ω 를 포함하고 있어서 직접 우도함수를 계산할 수 없으므로 그 분포를 가정하고 적분을 통하여 ω 에 의존하지 않는 우도함수(unconditional likelihood function) L 을 아래와 같이 구성하였다.⁴⁾

$$L = \sum_{k=1}^K [S_1(\bar{t}_1 | Z_1, \omega_k)]^{I_0} \left[-\frac{\partial S_1(t_1 | Z_1, \omega_k)}{\partial t_1} S_2(\bar{t}_2 | Z_2, \omega_k) \right]^{I_1} \\ \times \left[-\frac{\partial S_1(t_1 | Z_1, \omega_k)}{\partial t_1} \cdot -\frac{\partial S_2(t_2 | Z_2, \omega_k)}{\partial t_2} \right]^{I_2} \pi_k \quad (3)$$

여기서 \bar{t}_n 는 출산간격이 오른쪽으로 절단되어(right-censored) 완료되지 않은 기간을 의미하며, I_n 은 조사시점까지 n 번의 출산이 이루어졌음을 나타내는 더미변수이다. 식 (3)의 우도함수를 최대화하기 위하여 EM 알고리즘을 이용하였으며, 구체적 추정방법은 부록에서 보다 자세히 설명하였다.

IV. 자료

본 연구는 한국보건사회연구원에서 실시한 '2003년 전국 출산력 및 가족보건·복지실태 조사'의 자료를 사용하였다. 자료는 전국에서 표본추출한 200개 조사지역의 15~49세 기혼부인을 대상으로 임신, 출산 및 가족에 대한 질문에 대한 응답 내용을 포함하고 있는데, 출산 및 가구특성에 대하여 비교적 상세한 자료를 제공한다. 또한 임금 및 소득 방정식을 추정하기 위해 노동패널 자료를 이용하였는데, 노동패널 조사가 1998년부터 시작되

4) Heckman and Singer(1982)는 관찰되지 않는 이질성의 분포에 대한 자의적 가정에 따라 듀레이션 모형의 추정결과가 바뀔 수 있음을 지적하고, 유한한 반침(support)을 갖는 분포를 가정함으로써 이러한 위험에 덜 민감한 추정 결과를 얻을 수 있음을 보여주었다. 마찬가지로 본 연구에서도 ω 의 분포가 K 개의 반침 $\{\omega_1, \omega_2, \dots, \omega_k\}$ 과 이에 해당하는 확률 $\{p_1, p_2, \dots, p_k\}$ 를 갖는다고 가정하였다.

었으므로 추정된 임금 방정식이 유효하도록 출산력 자료 중 1998년 이후에 혼인한 여성으로만 분석 대상을 한정하였다. 연구에서 사용한 변수에 대해 무응답으로 처리된 경우 표본에서 제외하였으며, 결혼보다 먼저 출산을 하거나, 결혼 후 7개월 이내에 출산이 이루어진 경우는 결혼 이후에야 출산의 위험이 시작된다는 모형의 가정과 일관되지 않으므로 표본에서 제외하였다. 최종적으로 모두 822명의 여성이 표본에 포함되었다.

표본의 출산간격은 오른쪽으로 절단되므로 출산간격에 대한 요약 통계를 효과적으로 살펴보기 위해서 <표 1>에서는 조사시점까지의 총 출산회수에 따라 0회, 1회, 2회 이상의 세 그룹으로 구분하였다. 3명 이상의 자녀를 출산한 여성은 5명밖에 포함되지 않으므로, 2회 이상을 하나의 범주로 구분하였다. 조사시점까지 한 번도 출산을 하지 않은 여성의 경우 첫 출산간격은 오른쪽으로 절단되어 알 수 없으나, 첫 출산이 조사시점 이후에야 이루어질 것이므로 절단된 간격보다 더 큰 값을 갖는 것은 알 수 있다. 마찬가지로 조사시점까지 1번만 출산한 여성의 경우 첫 출산간격은 정확히 관찰되지만, 두 번째 출산간격은 절단되었다. 조사시점까지 2번 이상 출산한 여성은 첫 번째 출산간격과 두 번째 출산간격이 모두 정확히 측정되었다. 절단되지 않은 출산간격만을 살펴보면, 첫 출산간격은 평균 16.6개월, 두 번째 출산간격은 평균 24.9개월로 나타났다.

<표 1> 출산간격의 요약 통계 (단위: 개월)

조사시점까지 출산회수	첫 번째 출산간격	두 번째 출산간격	관찰값의 수
0	19.8 ^a (15.5)	-	281
1	18.0 (9.4)	24.3 ^a (14.1)	383
2+	13.2 (5.2)	24.9 (8.3)	158

a : 조사시점 현재까지 출산이 완료되지 않아서 오른쪽으로 절단된 값

주: 괄호 안은 표준편차

혼인연령 및 출산연령에 대한 통계를 살펴보면, 평균 혼인연령은 여성 배우자 27.0세, 남성 배우자 29.6세로 2.6세 정도의 차이가 나타난다. 또한 실제로 출산이 이루어진 경우에만 출산연령을 관찰할 수 있으므로 출산연령에 대한 관찰값의 수가 전체 표본수인 822보다 작음을 알 수 있다.

<표 2> 혼인 및 출산연령의 요약 통계 (단위:만 세)

	관찰값의 수	평균	표준편차	최소값	최대값
결혼연령(여성배우자)	822	27.0	3.2	18.3	47.5
결혼연령(남성배우자)	822	29.6	3.5	19.3	53.5
첫출산연령(여성배우자)	541	28.1	3.1	19.9	42.5
첫출산연령(남성배우자)	541	30.7	3.4	20.0	45.7
두번째출산연령(여성배우자)	158	29.7	2.9	22.4	39.4
두번째출산연령(남성배우자)	158	32.6	3.3	23.3	42.4

마지막으로 노동패널로부터 가져온 여성 임금과 남성 소득에 대한 요약 통계량을 <표 3>에 제시하였다. 본 연구에서 여성 임금은 여성 임금근로자의 주당 평균실질임금을 의미한다. 실제 노동패널조사는 임금근로자의 월평균임금과 주평균근무시간에 대한 항목을 제공하고 있으므로, (월평균임금*12)/(주평균근무시간*52)을 계산하고 이를 2000년 소비자물가지수로 조정하여 시간당 평균실질임금을 구하였다. 임금이 관찰된 여성은 모두 10,067명으로 시간당 임금의 평균값은 9,140원 정도로 나타났다. 반면 남성의 소득은 월평균 실질소득으로 정의되며, 임금근로자의 경우 월평균임금 항목을, 비임금근로자의 경우 월평균소득 항목을 사용하여 구하였다. 월평균소득이 관찰된 남성의 수는 모두 24,982명으로 월소득의 평균값은 114만원으로 나타났다. 남성의 월평균소득도 여성 임금과 마찬가지로 2000년 소비자물가지수로 조정하였다.

<표 3> 임금 및 소득의 요약 통계 (단위: 만원)

	관찰값의 수	평균	표준편차	최소값	최대값
여성 임금(주당평균임금)	10067	0.914	1.762	0.003	52.116
남성 소득(월평균소득)	24982	113.970	127.496	0.872	6726.475

V. 실증분석 결과

실증 분석은 여성 임금과 남성 소득을 설명하는 방정식을 추정하는 단계와 다음으로 이를 이용하여 출산시점에 관한 듀레이션 모형을 추정하는 단계로 나누어진다. 이렇게 두 단계로 나누어서 모형을 추정하는 이유는

첫 번째로 출산력 자료가 출산시점의 임금 및 소득 자료를 제공하지 않기 때문이고, 두 번째로 여성 임금이 근로자의 경우에만 관찰되므로 비근로 여성이 만약 근로할 경우 시장에서 받을 수 있는 임금 수준을 알아내기 위해서는 별도의 임금 방정식을 필요로 하기 때문이다. <표 4>는 노동패널 자료를 이용하여 여성 임금과 남성 소득을 설명하는 방정식을 추정한 결과이다. 첫 번째 열은 남성 월평균소득의 로그값을 연령, 연령²/100, 그리고 교육수준 더미변수에 회귀분석하여 소득 방정식을 추정한 결과이다. 추정계수의 크기에 비하여 표준오차가 작아서 모든 추정계수가 유의미하게 0과 다른 값을 갖는다. 연령에 대한 추정계수 0.247과 연령²/100에 대한 추정 계수 -0.279가 의미하는 바는 남성의 소득이 44.3세 까지 증가하다가 이후 감소함을 의미한다. 교육수준에 따라 남성의 월소득에 큰 차이가 나타났는데, 고등학교 졸업자는 중학교 졸업자에 비하여 월소득이 평균 20.1% 높았으며, 전문대 졸업자와 대학교 졸업 이상은 중학교 졸업 이하에 비하여 각각 29.3%와 38.3% 높은 소득수준을 갖는 것으로 나타났다.

<표 4>의 두 번째 열에는 여성의 임금 방정식을 추정한 결과를 제시하였는데, 여성의 임금 방정식과 임금노동시장 참여에 관한 프로비트(probit) 모형을 동시에 최우추정법에 의하여 추정한 결과이다. 노동시장 참여를 결정하는 설명변수로는 임금 방정식에 포함된 설명변수와 더불어 여성의 혼인여부를 나타내는 더미변수, 동거 중인 1세 이하 자녀수, 3세 이하 자녀수, 5세 이하 자녀수 및 18세 이하 자녀수를 이용하였다. 남성의 소득 방정식과 마찬가지로 모든 추정계수가 통계적으로 유의한 것으로 나타났으며, 교육수준이 높아짐에 따라서 시간당 임금도 상승함을 확인할 수 있다. 임금방정식과 임금노동시장 참여에 대한 선택 방정식의 오차항 간의 상관관계를 나타내는 ρ 는 0.004로 통계적으로 유의하지 않을 뿐 아니라, 두 오차항이 독립이라는 가설에 대한 우도비검정에서도 카이제곱 검정통계량의 p-value가 0.951로서 귀무가설을 기각할 수 없는 것으로 나타났다. 따라서 임금방정식의 추정에서 표본 선택으로 인한 문제는 심각하지 않은 것으로 볼 수 있다.

<표 4> 임금 및 소득 방정식

	ln(남성소득)	ln(여성임금)
연령	0.247 (0.005)	0.037 (0.010)
연령 ² /100	-0.279 (0.005)	-0.043 (0.013)
고등학교 졸업	0.201 (0.030)	0.230 (0.044)
전문대 졸업	0.293 (0.047)	0.403 (0.069)
대학교 졸업	0.383 (0.035)	0.691 (0.063)
상수항	-1.351 (0.104)	-1.972 (0.250)
ρ		0.004 (0.067)
관찰값의 수	24982	35854
R ²	0.112	
로그 우도함수값		-0.949

주:괄호 안은 표준오차

다음은 실증분석의 두 번째 단계인 듀레이션 모형의 추정결과로서 <표 5>에 보고하였다. 설명변수 중 ln(배우자소득)과 ln(여성임금)은 앞에서 추정한 소득 및 임금 방정식을 이용하여 구한 값을 사용하였으며, 그 외의 설명변수로는 대도시 더미 변수, 첫째 딸 더미 변수, 여성배우자의 결혼연령 및 첫 출산연령, 여성배우자의 출생연도에 따른 코호트 더미 변수가 포함되었다.⁵⁾ <표 5>의 첫 번째 열(A)에 제시된 결과는 설명변수 중 결혼연령과 첫 출산연령을 제외한 모형의 추정을 통해서 얻어진 것이며, 두 번째 열(B)은 연령변수를 포함한 모형의 추정 결과이다.⁶⁾ 두 모형의 추정결과에 큰 차이가 없으므로 아래에서는 연령변수를 포함한 모형의 추

5) 부모와의 동거여부는 우리나라의 상황에서 출산 결정에 영향을 미치는 매우 중요한 요인일 수 있으며 이를 통제할 필요성이 제기된다. 출산력 자료에서는 조사시점에서 부모와의 동거여부에 관한 정보를 포함하고 있으나, 출산 당시의 시점에 관한 자료는 제공하고 있지 않아서 본 연구에서는 설명변수로 포함하지 못하였다.

6) ω 의 분포는 K=5를 가정하였으며, 모형 B에 대해서 얻어진 ω 와 π 의 값은 아래 표와 같다. 위험함수에 대한 상수항을 별도로 추정하였으므로 ω 의 서로 다른 5개의 값을 모두 추정할 수는 없으므로 마지막 ω 의 값은 0으로 정규화 하였다.

ω	-1.504	-0.626	0.046	-4.108	0.000
π	0.263	0.343	0.088	0.041	0.265

정 결과를 중심으로 설명하였다.

$\ln(\text{여성임금})$ 에 대한 계수는 첫 출산의 경우 -0.959 으로 나타났으며 표준오차는 0.224 로서 통계적으로도 유의하게 0보다 작다고 볼 수 있다. 이는 여성 임금의 상승으로 인한 기회비용의 증가가 첫 출산을 유의미하게 지연시키고 있음을 나타낸다는 것을 확인한 것으로서, Heckman and Walker(1990), Merrigan and St.-Pierre(1998), 류덕현(2007)의 결론과 비슷하다. 또한 추정된 계수를 정량적으로 해석하기 위해서 식 (1)을 살펴보면, 위험률의 로그값과 여성 임금의 로그값이 선형관계를 갖는 것으로 상정하였으므로 추정된 계수가 탄력성을 의미함을 알 수 있다.⁷⁾ 즉, 추정결과는 여성 임금이 1% 증가하였을 때 출산위험률이 0.96% 감소한다는 것을 의미한다.

반면 두 번째 출산의 경우 $\ln(\text{여성임금})$ 의 계수는 -0.743 으로 첫 출산과 비교하여 계수의 절대적 크기가 감소하였을 뿐 아니라, 표준오차가 0.494 로서 통계적 유의성도 가지지 못하는 것으로 나타났다. 즉 이러한 결과는 여성 임금의 상승은 첫 출산을 지연시키지만 그 이후의 출산에 대해서는 이러한 효과가 나타나지 않을 수 있음을 의미한다. 출산 순서에 따라 출산지연의 효과가 다르게 나타날 수 있다는 점은 흥미로운 결과로서, 출산장려정책의 효과성을 높일 수 있는 지원 대상 선정에 대하여 시사점을 가질 수도 있다.

$\ln(\text{배우자소득})$ 에 대한 계수는 첫 출산의 경우 0.164 , 표준오차는 0.187 로 나타나서 통계적으로도 유의미하게 0보다 크다고 볼 수 없을 뿐 아니라, 절대적 크기에 있어서도 여성 임금의 효과보다 작다. 첫 출산시점의 결정과 관련하여 배우자 소득의 증가로 인한 효과가 여성 임금의 효과에 비해 약하다는 결론은 Heckman and Walker(1990)에서 발견한 결과와 일치한다. 한편 배우자 소득이 두 번째 출산에 대해 미치는 효과는 그 추정계수가 1.020 으로 첫 출산에 비하여 높은 값을 가질 뿐 아니라, 통계적으로 유의한 것으로 나타났다. 즉 배우자 소득의 증가로 인해 출산이 앞당겨지는 효과는 첫 출산에 대해서는 약한 반면, 그 이후 출산에 대해서는 강하게 나타날 수 있다는 것이다.

또한 모형에서는 여성 배우자의 출생연도에 따라 코호트를 정의하고,

7) 확률에 대한 탄력성이 비율에 대한 비율이라서 혼동이 올 수 있다. 예를 들어 현재 위험률이 50%인 경우 여성 임금의 1% 증가는 위험률을 0.5%p 낮추게 된다는 의미이다.

이에 해당하는 더미변수를 포함시켜서 출생연도에 따른 출산 패턴의 변화가 반영되도록 하였다. 표본의 여성을 1960년생부터 5년 간격으로 구분하여 1960~64년생, 1965~69년생, 1970~74년생, 그리고 1975년 이후 출생자의 네 출생코호트로 나누고, 이 중 앞의 세 코호트에 해당하는 더미변수를 모형에 포함하였다.⁸⁾ 따라서 각 더미 변수에 대한 계수의 추정값은 1975년 이후 출생자에 대비한 출산위험율의 차이를 의미한다. 첫 출산의 경우 출생코호트 더미에 대한 계수는 0.755, 0.561, 0.244로 나타나서 최근에 태어난 여성일수록 출산을 지연하고 있음을 나타내고 있으나, 각 계수의 추정치가 통계적으로 유의한 수준에 이르지 못하는 못하였다. 두 번째 출산의 경우에도 대체로 유사한 패턴이 발견되었다.

여성 임금, 배우자 소득, 출생코호트 이외에 기타 추정치에 대해서 간략하게 설명하면 다음과 같다. μ 는 각 출산 단계에서의 출산 중단 확률을 측정하는 파라미터로서, 각 추정치가 의미하는 출산 중단 확률 p 를 <표 5>에 별도로 표시하였다. 추정 결과에 따르면 혼인 이후 출산을 중단한 여성의 비율은 8.9%, 첫 번째 출산 이후 두 번째 출산을 중단한 여성의 비율이 39.7%인 것으로 나타났다. γ 는 첫 번째 출산과 두 번째 출산 모두에 대해서 양의 값을 나타내는 것으로 추정되어서, 시간이 지남에 따라서 출산의 확률이 점차 높아짐을 의미한다. 첫자녀 딸 더미변수는 전통적인 남아선호로 인하여 출산행동이 바뀔 가능성을 반영하기 위한 변수이다. 만약 부모의 아들에 대한 선호가 매우 강하다면, 첫 자녀가 딸인 경우 두 번째 출산의 가능성이 높아질 것이다. 추정결과는 0.187로 첫 자녀의 성별의 효과가 통계적으로 유의미하지 않은 것으로 나타났는데, 이는 최근 들어 아들에 대한 선호가 뚜렷하지 않다는 것을 시사한다. 결혼연령이 첫 출산시점에 미친 효과와 첫 번째 출산 연령이 두 번째 출산시점에 미치는 효과는 모두 통계적으로 유의하지 않았다.

8) 표본에서 각 코호트의 비중은 1960~64년생은 2.1%, 1965~69년생이 10.1%, 1970~74년생은 51.2%, 그리고 1975년 이후 출생자는 36.6%이다.

<표 5> 듀레이션 모형의 추정 결과

	(1)	(2)
	첫 번째 출산	
δ_1	-1.848 (0.957)	-0.733 (1.393)
γ_1	1.594 (0.075)	1.571 (0.076)
ln (배우자소득)	0.106 (0.184)	0.164 (0.187)
ln (여성임금)	-1.048 (0.210)	-0.959 (0.224)
대도시 더미변수	-0.074 (0.085)	-0.086 (0.085)
결혼연령(여성배우자)		0.034 (0.028)
더미(1960~1964)	0.973 (0.460)	0.755 (0.496)
더미(1965~1969)	0.901 (0.455)	0.561 (0.544)
더미(1970~1974)	0.666 (0.462)	0.244 (0.596)
μ_1	-2.290 (0.153)	-2.328 (0.158)
출산중단 확률 (p_1)	0.092	0.089
	두 번째 출산	
δ_2	-8.169 (2.019)	-8.972 (3.068)
γ_2	2.641 (0.221)	2.668 (0.239)
ln (배우자소득)	1.071 (0.396)	1.020 (0.429)
ln (여성임금)	-0.746 (0.491)	-0.743 (0.494)
대도시 더미변수	0.152 (0.197)	0.160 (0.198)
첫자녀 딸 더미변수	0.178 (0.179)	0.187 (0.180)
첫출산연령(여성배우자)		0.027 (0.078)
더미(1960~1964)	1.839 (0.700)	2.001 (0.850)
더미(1965~1969)	1.355 (0.680)	1.597 (0.986)
더미(1970~1974)	1.540 (0.743)	1.871 (1.210)
μ_2	-0.438 (0.151)	-0.417 (0.160)
출산중단 확률(p_2)	0.392	0.397
관찰값의 수	822	822
로그우도합수값	-1.532	-1.533

주: 괄호 안은 표준오차

마지막으로 [그림 2]와 [그림 3]에서는 추정모형의 가장 중요한 변수인 여성 임금과 배우자 소득의 효과를 보다 쉽게 확인할 수 있도록 추정된 모형이 의미하는 출산의 생존함수와 그 변화를 그래프로 나타내었다. 식 (4)에서와 같이 각 여성의 생존함수값을 구한 후 전체 표본에 대해 평균을 계산하여, 평균적인 생존함수 \bar{S}_j 를 구하였다.

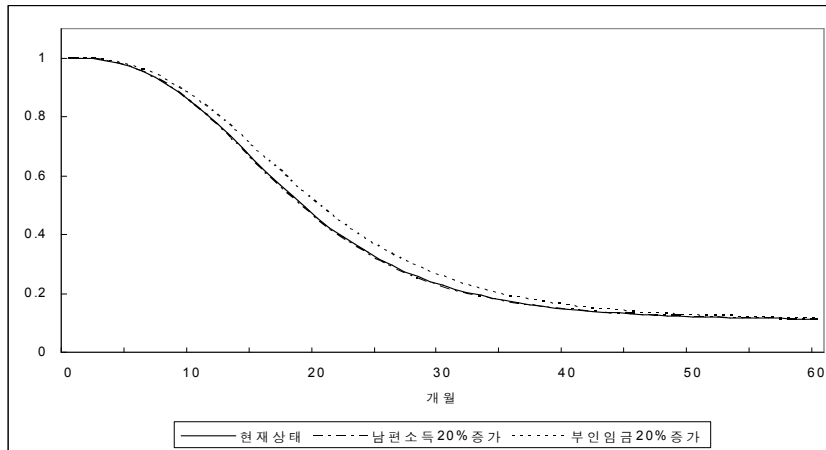
$$\bar{S}_j = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K S_{ji}(t|Z_{ji}, \omega_k) \pi_k, \quad j = 1 \text{ 또는 } 2 \quad (4)$$

[그림 2]에서 ‘현재상태’로 표시된 실선이 추정 모형이 의미하는 첫 출산의 생존함수이며 각 시점에서 첫 출산을 완료하지 않은 여성의 비율이다. 생존함수가 감소하는 것은 혼인 이후 시간이 지남에 따라 점점 많은 비율의 여성이 첫 출산을 완료하는 것을 의미한다. 감소하던 생존함수가 일정 시간 이후 더 이상 감소하지 않고 거의 일정한 수준을 유지하는 것을 알 수 있는데, 바로 이 수준이 첫 출산을 아예 하지 않는 비율로서 위에서 설명한 출산중단 확률에 해당한다. 생존함수가 50%에 도달하는데 소요되는 기간은 19개월로서, 혼인 이후 19개월에 표본의 여성 중 50% 정도가 첫 출산을 완료한다는 의미이다. 생존함수가 20%에 도달하는 시점, 즉 80%의 여성이 출산을 완료하는 시점은 혼인 이후 33개월인 것으로 나타났다. 또한 [그림 2]의 두 점선은 남성 소득이 20% 증가한 경우와 여성 임금이 20% 상승한 경우를 상정하여 각각의 경우 첫 출산의 생존함수를 표시한 것이다. 남성소득의 증가는 효과가 미미하여 원래 생존함수와 구별하는 것이 거의 힘든 반면, 여성 임금의 20% 증가는 생존함수를 오른쪽으로 이동시켜서 첫 출산이 그만큼 지연되었음을 의미한다. 여성임금이 20% 증가한 경우 생존함수가 50% 수준에 도달하는데 걸리는 시간은 21개월로서 임금이 증가되지 않은 상태와 비교하여 2개월 늦어졌으며, 20%에 도달하는데 걸리는 시간도 35개월로서 임금 증가 이전에 비하여 2개월 지연되었다.

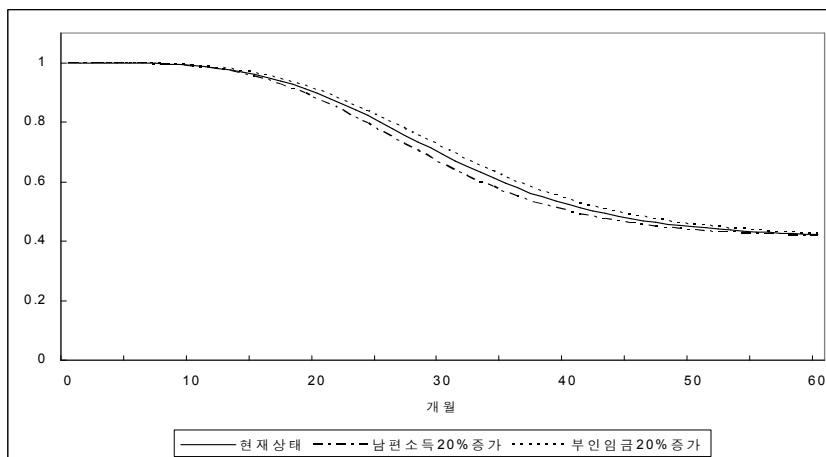
[그림 3]에서는 두 번째 출산에 대한 생존함수를 [그림 2]와 마찬가지로 나타내었다. 두 번째 출산의 경우 생존함수가 수렴하는 수준이 높아져서 40% 정도에서 평평한 구간을 이루고 있음을 알 수 있다. 여성 임금의 증가는 첫 출산에서와 마찬가지로 생존함수를 오른쪽으로 이동시켜 출

산이 지연될 것을 예측하며, 배우자소득 증가의 효과는 첫 출산에 비하여 비교적 뚜렷하게 나타나서 생존함수가 왼쪽으로 이동한 것을 확인할 수 있다. 생존함수가 50% 수준에 도달하는데 소요되는 기간은 현재의 임금 및 소득 수준에서는 43개월, 배우자소득이 20% 증가한 경우는 41개월, 여성임금이 20% 증가한 경우는 44개월로 각각 나타났다.

[그림 2] 첫 번째 출산 간격에 대한 생존함수



[그림 3] 두 번째 출산에 대한 생존함수



VI. 요약 및 결론

출산율의 변화는 여성이 일생에 걸쳐 출산하는 자녀 수와 더불어 출산 시점의 선택에 의하여 큰 영향을 받게 된다. 본 연구는 여성의 임금과 배우자의 소득이 첫 번째 및 두 번째 출산간격에 미치는 효과를 실증적으로 검토하고 그 효과의 크기를 분석하였다. 연구결과를 간략하게 요약하면, 여성임금의 효과는 첫 번째 출산을 유의미하게 지연시키는 효과를 갖는 것으로 나타났으며, 출산 위험율의 여성 임금에 대한 탄력성은 -1 정도로서 여성 임금의 1% 증가는 출산 위험율을 1% 정도 감소시킬 것으로 예측되었다. 반면 두 번째 출산간격에 대한 여성 임금의 효과는 첫 출산에 비하여 다소 약한 것으로 나타났다. 표본의 여성을 대상으로 모형이 의미하는 효과의 크기를 살펴보면, 여성 임금의 20% 증가에 대하여 표본 중 50%가 출산을 완료하는 시점을 기준으로 첫 출산은 약 2개월, 두 번째 출산은 약 1개월 정도 늦어지는 것으로 나타났다. 한편 배우자 소득이 출산을 앞당기는 효과는 두 번째 출산에 대해서만 통계적 유의성이 발견되었다.

본 연구결과를 바탕으로 향후 두 가지 측면에서 추가적 연구의 필요성이 제기된다. 먼저 장래 출산율을 예측하기 위해서 기혼자의 출산결정과 더불어 미혼자의 결혼 결정을 모형에 포함할 필요가 있다. 특히 최근 출산 활동 위축의 상당부분은 결혼이 늦어지는데서 비롯된 것으로 의심되기 때문에, 미혼자가 결혼을 미루는 현상을 포함하지 않는다면 장래의 출산율을 체계적으로 예측하는데 어려움이 있다. 다음으로는 정책의 효과를 평가할 수 있도록 여성의 출산 결정과 노동공급을 동시에 고려하는 모형을 추정하는 것이다. 이미 살펴보았듯이 여성의 노동시장 참여는 출산을 저해하는 요인이 되는데, 정책 수단을 통하여 과연 여성 고용을 희생하지 않고 출산을 장려할 수 있는지 여부는 매우 중요한 정책적 관심사항이다. 이 문제에 대한 연구는 추후의 연구과제로 남기고자 한다.

[참고문헌]

류덕현, 「출산율과 여성노동공급에 대한 거시적 실증분석」, □ 공공경제□, 제 12권 제1호, 한국재정학회, 2007, 39-74.

- 은기수, 「결혼연령 및 결혼코호트와 첫 출산간격의 관계」, □ 한국사회학□ , 제35집 6호, 2001, 105-139.
- 이삼식 외, □ 저출산 원인 및 종합대책 연구□ , 연구보고서 2005-30(2), 한국보건사회연구원, 2005.
- Becker, G., “An Economic Analysis of Fertility”, *Demographic and Economic Change in Developed Countries*, Conference of the Universities-National Bureau Committee for Economic Research, 1960, 209-240.
- Heckman, J. and B. Singer, “The Identification Problem in Econometric Models of Longitudinal Data”, W. Hildenbrand (ed). *Advances in Econometrics*, 1982. 39-77.
- Heckman, J. and J. Walker, “Relationship between Wages and Income and the Timing and Spacing of Births: Evidence from Swedish Longitudinal Data”, *Econometrica* Vol. 58, No. 6, 1990, 1411-1441.
- Klasen, S. and A. Launov, “Analysis of the determinants of fertility decline in the Czech Republic”, *Journal of Population Economics*, Vol. 19, No. 1, 2006, 25-54.
- Lancaster, T., *The Econometric Analysis of Transition Data*, 1992.
- Merrigan, P. and Y. St.-Pierre, “An econometric and neoclassical analysis of the timing and spacing of births in Canada from 1950 to 1990”, *Journal of Population Economics*, Vol. 11, No. 1, 1998, 29-51.
- OECD, *Society at a Glance: OECD Social Indicators - 2006 Edition*, 2006.
- Tasiran, A., *Fertility Dynamics: Spacing and Timing of Births in Sweden and the United States*, 1995.

<부록> EM 알고리즘을 이용한 듀레이션 모형의 추정

ω 의 분포가 K개의 유한한 받침을 갖는다고 가정하였으나, 받침의 수의 실제값은 알려져 있지 않다. 만약 K가 실제 받침의 수보다 클 경우에는 우도함수가 파라미터 공간의 일정한 선(ridge)을 따라서 최대값을 가지는 문제점이 발생한다. EM 알고리즘은 전체 파라미터에 대한 헤시안 행렬을 구하지 않아도 되기 때문에, 연구자가 관심을 가지는 주요한 파라미터에 대한 최우추정을 용이하게 만든다.

EM 알고리즘은 반복적으로 파라미터를 업데이트하는 다음의 과정을 통하여 우도함수를 최대화시키는 모수값을 찾는다. 먼저 m-1단계의 파라미터를 이용하여 구한 로그우도함수 $\ln L_{ik}^{(m-1)}$ 와 $\Pi_k^{(m-1)}$ 를 이용하여 아래와 같이 $\pi_{ik}^{(m)}$ 을 구하는 과정으로 이를 E스텝(expectation step)이라고 부른다.

$$\pi_{ik}^{(m)} = \frac{\Pi_k^{(m-1)} \ln L_{ik}^{(m-1)}}{\sum_{k=1}^K \Pi_k^{(m-1)} \ln L_{ik}^{(m-1)}} \tag{A1}$$

$$\Pi_k^{(m)} = \frac{1}{N} \sum_{i=1}^N \pi_{ik}^{(m)} \tag{A2}$$

여기서 구해진 $\pi_{ik}^{(m)}$ 는 주어진 m-1단계의 파라미터에 대해서 관찰값 i 가 ω_k 로부터 구해졌을 사후적 확률을 베이지 규칙을 이용하여 계산한 것이다. 다음 스텝은 M스텝(maximization step)으로 E스텝에서 구한 $\pi_{ik}^{(m)}$ 를 이용하여 다음 함수를 최대화하는 파라미터를 구한다.

$$\mathcal{LAPLACE}^{(m)} = \sum_{i=1}^N \sum_{k=1}^K \pi_{ik}^{(m)} \ln L_{ik} \tag{A3}$$

m이 증가하면서 E스텝과 M스텝을 반복함에 따라 본래 로그우도함수인

$$LAPLACE = \sum_{i=1}^N \ln \left(\sum_{k=1}^K L_{ik} \pi_{ik} \right) \quad (A4)$$

의 최우추정량을 구할 수 있음이 알려져 있다 (Heckman and Singer 1982; Lancaster 1992).

[Abstract]

Effects of Wages and Income on the Birth Intervals in Korea

Hee Chul Min

This paper empirically studies the effects of female wages and male income on the timing of births for 822 recently-married women in the “Survey on the National Fertility, Family Health and Welfare”. A duration model following Heckman and Walker(1990) was adopted to model the birth process, combined with the income and wage equations estimated from KLIPS (Korea Labor and Income Panel) data. The estimated model suggests that a rise in female wage significantly delays first births while the effect on second births is rather weak. A 20% increase of female wage is expected to delay first births by two month and second births by one month, as measured by the time it takes for the estimated survival functions to reach 50%. On the other hand, the effect of male income on the first birth interval is insignificant with relatively small magnitudes.

Keywords: fertility, timing of birth, duration model

JEL Code: J10, J13, C41