

제3장

장래인구추계 출산력 모형 연구

오진호

제1절 서론

1. 연구배경 및 목적

우리나라는 2016년 이후에 인구 오너스(Demographic Onus)¹⁾시대가 도래함에 따라 출산율 제고에 대한 관심과 정책이 대두되고 있다. 생산가능인구 비중이 감소하고 유소년층과 고령층 등 부양대상 인구가 증가하는 시기를 인구학적 오너스 시대라고 하는데 우리나라는 이러한 시기가 2016년 이후에 도래된다. 이에 대한 원인으로 저출산이 지배적인 의견인데 이에 대한 정확한 미래 추계가 필요하다.

우리나라 출산율은 인구대체출산율(replacement level fertility)²⁾ 2.1보다 현저히 아래에 있으며, 2005년 최저점인 1.076에서 2009년 이후 4년 연속 증가해 2012년 1.297을 보이다가 2013년 최근에 다소 주춤한 1.185를 나타내고 있다(<표 3-1> 참조). 우리나라는 인구감소 수준의 출산율이 적지 않은 기간 이어오고 있으며 세계 최고 수준의 고령화 속도와 맞물려 급속한 인구구조 변화를 겪고 있다.

저출산 문제가 현대사회에 제기하는 사회·경제적 함의는 다양하다(이문숙, 2014).

1) MaDonald, P.(2006). "Low Fertility and the State: The Efficacy of Policy." Population and Devopment Review 32(3), pp.401~510

2) 합계출산율이 2.1일 때를 인구대체출산율(replacement level fertility)이라고 하며, 1.5일 때를 인구대체 안전선(safty zone)이라 한다. 이 안전선은 미래의 노동력을 이민 등의 정책으로 보완할 수 있는 출산율이다 (McDonald, 2006).

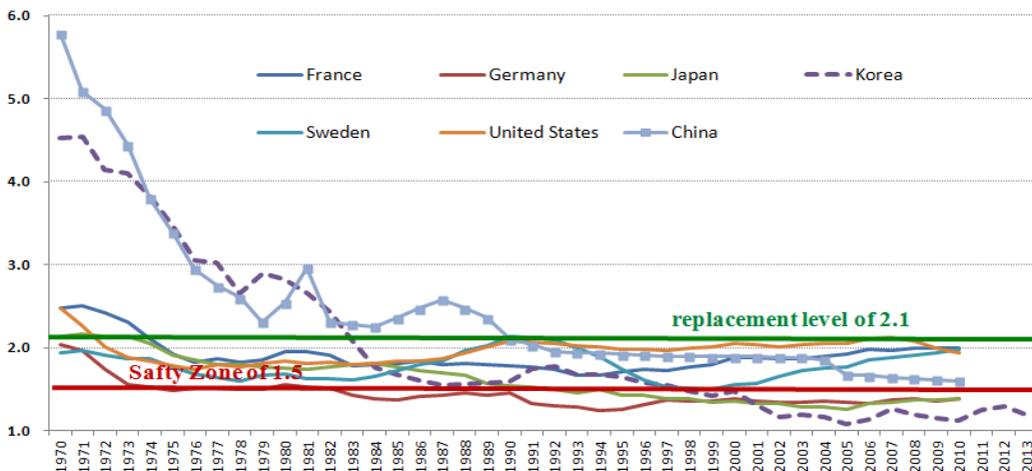


경제적인 측면에서 인구증가는 시장의 규모를 확대시킴으로서 전문화, 분업 등으로 규모의 경제를 가능하게 하고 기술발전을 이끌지만 저출산은 노동인력의 부족, 소비시장 약화, 내수시장 비활성화 등으로 경제성장에 부정적인 영향을 미친다. 사회적인 측면에서의 영향은 저출산으로 인한 가족안전망 해체이다. 가족관계가 없는 사람들의 숫자가 늘어나게 되어 정부 또는 민간분야의 보편적 복지에 대한 공급수요가 더욱 늘어나게 된다. 많은 수의 유아들이 형제나 친구가 없어져 세대간 유대관계에도 변화가 예상된다. 또한 많은 수의 활동적인(active) 노년층이 차지하고 있는 정부와 기업체의 지위에 젊은 세대의 진입이 견제 받게 됨으로써 세대간 갈등이 심화될 수 있다. 또한 저출산은 학령인구(6~21세)의 감소로 이어져 학교와 교원 수의 감축을 초래하게 된다.

〈표 3-1〉 최근 10년간 모의 연령별 출산율 및 합계출산율

연도	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
합계 출산율	1.180	1.154	1.076	1.123	1.250	1.192	1.149	1.226	1.244	1.297	1.185
연령별 출산율	15-19세	0.012	0.011	0.010	0.011	0.012	0.009	0.009	0.009	0.009	0.008
	20-24세	0.120	0.100	0.085	0.084	0.094	0.089	0.082	0.082	0.083	0.071
	25-29세	0.553	0.521	0.463	0.454	0.485	0.429	0.394	0.385	0.377	0.323
	30-34세	0.398	0.419	0.413	0.455	0.515	0.515	0.508	0.562	0.570	0.606
	35-39세	0.084	0.089	0.092	0.105	0.128	0.134	0.138	0.167	0.182	0.200
	40-44세	0.012	0.013	0.012	0.013	0.015	0.015	0.017	0.020	0.022	0.025
	45-49세	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

자료 : 통계청(2013)



[그림 3-1] OECD 대표 저출산 국가들과 출산율추이 비교

OECD 국가들과 출산율 추이를 비교해보면, 우리나라 경우 인구대체율과 인구안전선을 지나 2013년 1.18에 도달하는 모습이 세계에서 가장 빠르며³⁾, 효과적인 출산장려정책을 시행하지 않으면 출산율 저하가 반복되는 소위 Lutz et al.(2006)의 저출산 덩(the low fertility trap)에 빠질 가능성이 높다.

이런 현실을 감안할 때 올바른 사회정책이 펼쳐지기 위해 정확한 장래인구추계의 한 축인 출산율에 대한 정확한 예측모형에 대한 요구가 대두되고 있고 추계의 예측력 향상을 위해 체계적인 출산력 추계모형의 개선 및 개발이 필요하다.

그리고 인구규모 및 인구구조를 변동시키는 출생·사망·혼인·이혼 중 출생은 인구, 복지, 보건, 주택 등에 관한 정책수립 및 민간기업의 경영계획 수립에 필요한 기초 자료 제공으로 중요한 부분이며, 체계적인 출산력 모형 개선 및 개발을 위해 해외사례 및 최신동향 검토가 요구된다.

2. 연구내용 및 범위

본 연구에서는 장래인구추계의 출산력 예측의 최신동향을 모수화모형(parameterized model) 변천, 모형의 간편성을 강조한 시계열적합 및 예측, 예측값 활용 중심으로 살펴보고자 한다.

흔히 출산예측모형의 분류법은 많지만, 기본적으로 수리적·경험적 모형(mathematical empirical model), 코호트·기간 모형(cohort period model), 연령별 지속기간별 모형(age-duration-specific model) 등을 들 수 있다(전광희 2006). 이중 출산예측모형은 코호트의 출산행태에 입각한 파라미터 모형을 이상적으로 제안하고 있다. 즉 확률분포모형을 활용한 모수화모형이 우수하다고 알려져 있다. 이러한 모수화모형의 유용성을 논할 때는 과거 출산율 패턴이 미래에도 지속되어야 한다는 가정⁴⁾의 성립이 필요한데, 이러한 가정은 우리나라와 같이 최근까지 출산율이 감소추세인 경우에는 당연히 출산율이 감소할 것으로 예측되어 정부에서 선택적 복지로 펼치고 있는 출산장려정책이 반영되지 않는 결과가 도출된다.

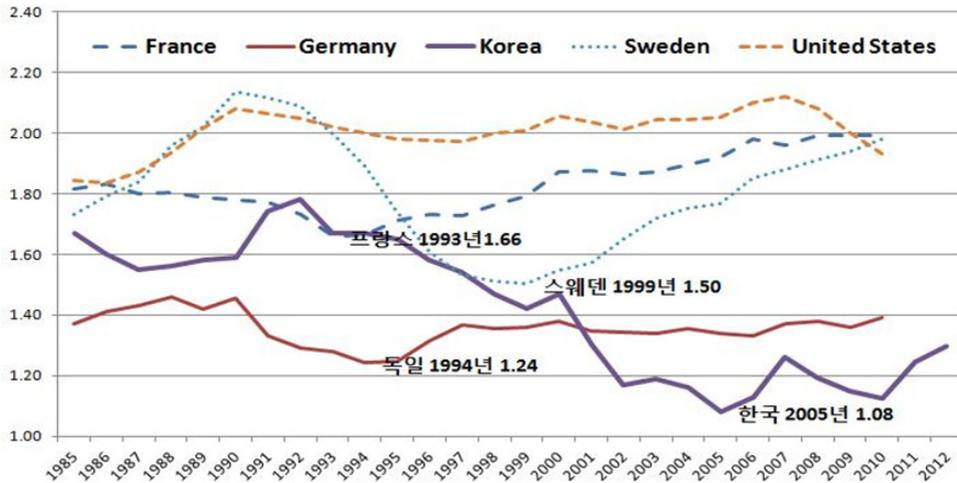
일반적으로 출산율 3단계 관점(높은 출산율 단계, 출산율 전이단계, 최저출산율 이후 단계⁵⁾)에서 최저출산율 이후단계에 도달한 선진국들의 합계출산율 추세를 보면 프랑스는

3) 9개국 출산율 비교결과 및 1.5 이하 국가인 독일, 프랑스, 중국, 일본을 비교하여 출산율 추이를 <부록>에 제시한다.

4) 이러한 가정은 최근 UN에서 사용한 Alkema et al.(2010) 방식의 2.1로 수립한다는 것과 일맥상통한 것이다.

5) 인구변천이론과 비슷한 의미로 제1의 인구변천은 출생률과 사망률이 높은 단계로 시작해 점진적 감소기조를 나타내는 것을 의미한다. 이후 인구대체를 불가능하게 하는 저출산 시대가 시작되고 고령사망률의 저하와 함께 기대수명이 지속적으로 증가하여 인생 100세라는 장수화 시대에 진입하였다. 저출산과 장수화는 인구의 급격한 고령화를 가져오고 인구감소의 시대가 시작되는데 이를 제2의 인구변천이라 한다.

1993년 1.661, 독일은 1994년 1.244를 기록하고 출산장려정책에 힘입어 점진적 상승을 보이고 있다.



[그림 3-2] 선진국 출산율과 한국 출산율 추이

본 연구에서는 연령별 출산율을 예측하기 위해 각세별 시계열적합과 모수화방법을 동시에 활용하고자 한다. 즉, 가임인구(15~49세)의 각세별 출산율 추이에 대해 시계열방법을 써서 연령별 시계열 모형을 구한 후 최적모형을 선별해 예측을 하고자 한다. 그런다음 예측결과를 모수화방법에 적용하여 출산순위별 출산율과 합계출산율을 제안하고자 한다.

이 방법은 모수화모형 예측의 가정에서 감소 기조의 출산율 추이를 반영하지 못한다는 단점을 개선할 수 있다.

본 연구의 구성은 다음과 같다.

2절에서는 첫째 연령별 출산율을 확률과정으로 가정해 정상·비정상 시계열 판정 후 시계열 적합, 진단, 예측의 3단계를 거쳐 2014~2060년의 연령별 출산율을 예측한다.

둘째 출산율 예측을 위한 모수화모형에 대한 내용을 다룬다. 출산율 예측모형의 변천사, 선진국과 다른 여러나라들에 적합한 출산율 모형과 최신동향 살펴본다. 그리고 출산율 예측모형의 변천 특징을 5가지로 파악해 5개의 출산율 예측모형을 제시해 본다.

셋째, 연령별 출산율 시계열 예측결과를 제안된 모수화모형에 대입하여 최종적인 2014~2060년의 출산율 추이를 알아본다. 그리고 통계청(2010)의 장래인구추계결과와 비교하여 적합성을 논한다.

3절에서는 결론과 연구결과의 시사점을 도출하기로 한다.

제2절 출산율 적합과 예측

연령별 출산율을 직접 추계하는 방법으로 첫째 수리적 모형(수정지수 방법, 로지스틱 방법 등)으로 추세를 연장하는 방법, 둘째 연령별 출산율과 사회경제적 변수와의 관계를 통해 추계하는 방법, 셋째 구미 선진국의 연령별 출산율 동향이나 목표치를 고려하여 장래 추정 시점의 연령별 출산율을 가정한 후 선형방법으로 연장하는 방법 등이 있다(통계청 2013).

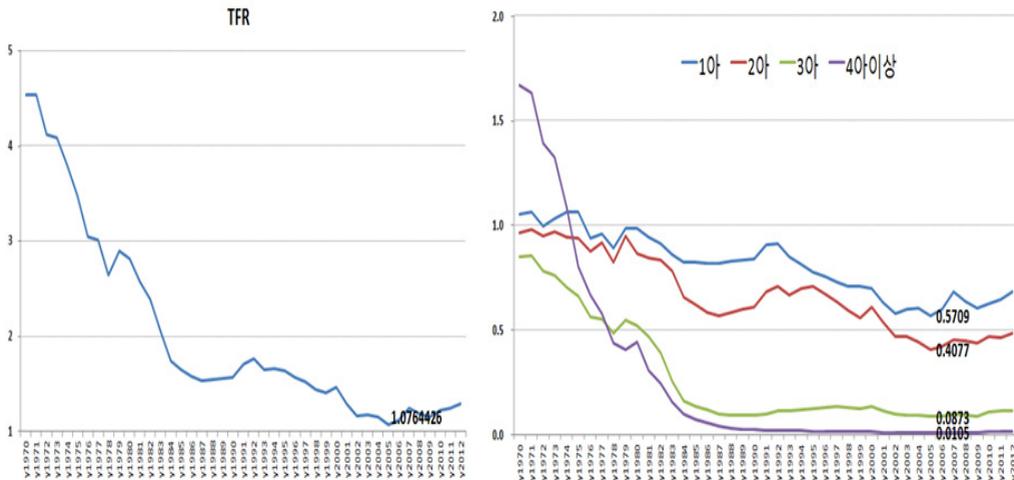
앞절에서 제시한 바와 같이 출산율의 적합과 예측을 위해 가장 널리 사용되는 방법은 모수화모형이다. 관측된 연령별 출산율을 다양한 함수에 적용하여 모수를 추정된 후 이 추정된 모수를 시계열 모형에 적합하여 미래의 출산율을 예측한다. 하지만 이러한 모수화 모형의 유용성은 과거 출산율 패턴이 미래에도 지속되어야 한다는 가정이 성립되어야 가능하다는 것을 앞절에서 언급하였다.

본 연구에서는 첫째 방법과 유사점이 있지만 약간 다르게 시계열방법을 연령별 출산율에 적용하여 가임인구(15~49세)에 대해 각각 시계열 추이를 분석해 적합 예측하는 방법을 적용한다.

그리고 이 결과를 모수화모형에 적용하여 예측치를 도출한다. 이러한 방법은 모수화 모형의 출산율 지속가능성의 가정을 보완할 수 있다.

1. 연령별 출산율 시계열분석

우리나라의 합계출산율은 1970년 4.53 수준에서 1983년 인구대체율인 2.1 이하인 2.08, 2000년에 안전선(safety zone) 아래인 1.47, 2005년 1.08로 최저점을 찍고, 2011년



[그림 3-3] 출산율 추이

1.244, 2012년 1.297로 높아지는 듯했으나 2013년 1.187로 낮은 수준에 머물러 있다. 출산율이 낮아진 주된 이유는 사회경제적 환경 및 가치관 변화에 따라 결혼의 기피와 연기로 인한 결혼연령 상승과 출산지연 때문인 것으로 판단된다.

세부적으로 모의 연령별 출산율(Age-specific Fertility Rate; ASFR)을 알아보자. ASFR은 모의 각세별 출생수를 해당연령별 인구로 나누어 출산율을 계산한다. 여성 1명이 가임 기간 동안 낳을 것으로 예상되는 평균 출생수를 합계출산율(Total Fertility Rate; TFR)이라고 한다. 다음 [그림 3-4]에서 ②에 해당되는 부분이다.

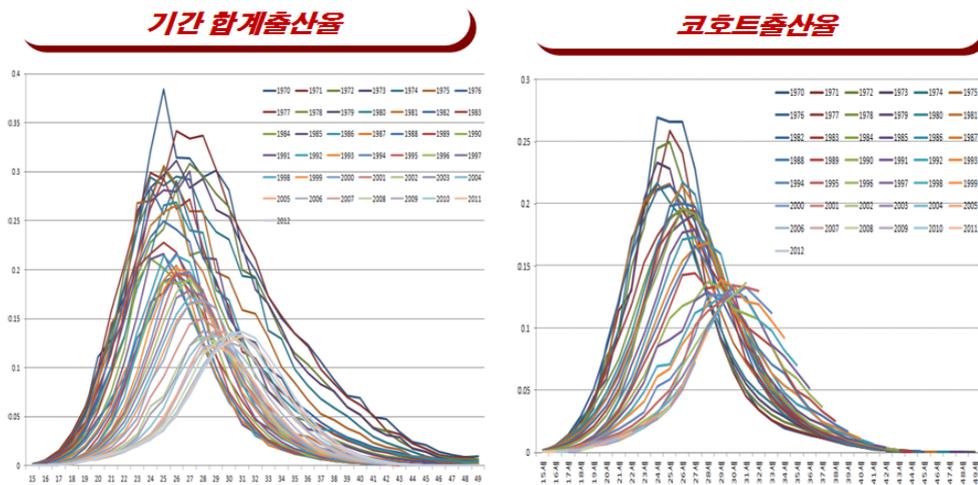
연령별 출산율을 도출함에 있어 공식적으로 장래인구추계에서는 코호트출산율(Cohort fertility) 방법이 널리 쓰이고 있다. 코호트출산율에 의한 추계방법은 여자의 연령별 출산율을 코호트로서 관찰한 것으로 코호트 완결 출생아수와 연령별 출생률을 추계하는 방법이다. 코호트 방법은 [그림 3-4]의 ①에 해당된다. 코호트로 관찰하는 것으로 15세 t년, 16세 t+1년, 17세 t+2년을 살펴보는 방법이다.

연령	t년	t+1년	t+2년	t+3년
15	f(15)			
16	f(16)	f(16)		
17	f(17)		f(17)	
18	f(18)			f(18)
...				

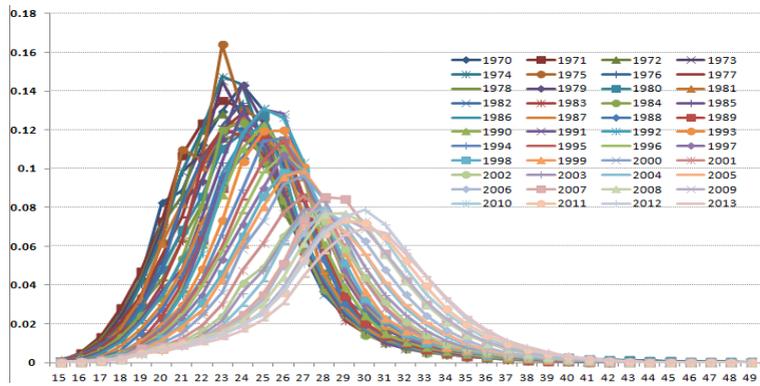
②

①

[그림 3-4] 모의 연령별 출산율과 코호트 출산율

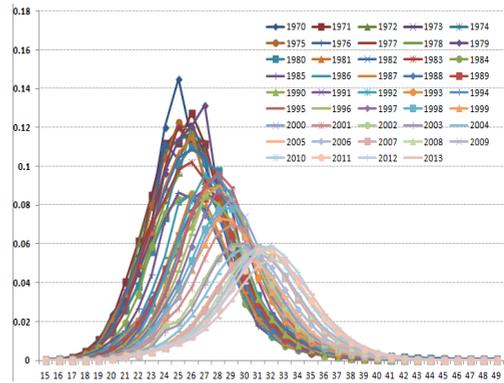
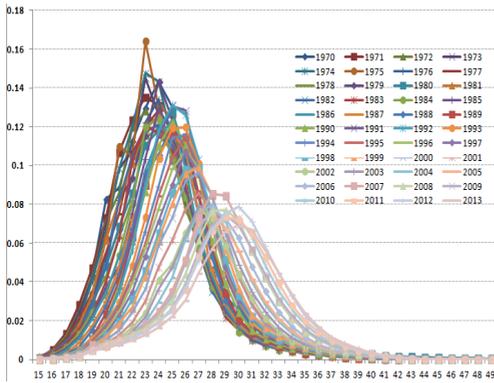


[그림 3-5] 기간 합계출산율과 코호트출산율 비교



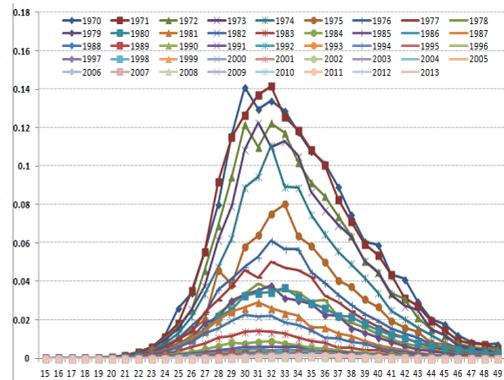
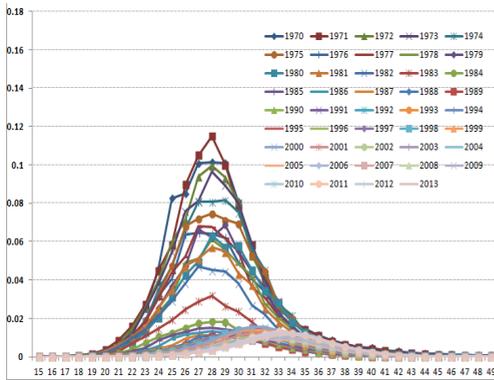
(10)

(20)



(30)

(40 이상)



[그림 3-6] 연령별 합계출산율(TFR)과 출산순위별 출산율



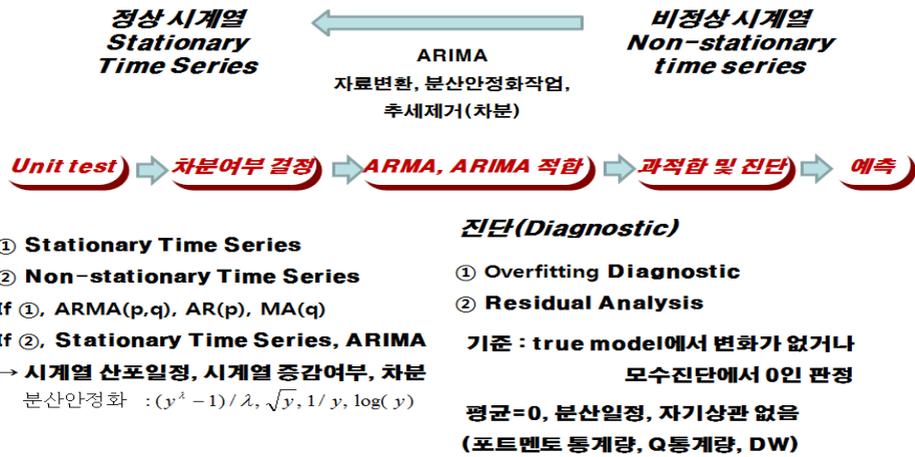
[그림 3-5]에서 알수 있듯이 연령별 코호트완결출산율을 산출하기 위해서는 젊은 출생코호트의 경우에 그 이전의 출생코호트를 대상으로 해서 구한 각각의 파라미터 값의 추이를 바탕으로 장래의 특정 출생 코호트에 대한 가정치의 값들을 예측하여야 한다.

이처럼 일생동안 출산하게 될 자녀수를 나타내는 코호트출산율은 기간출산율보다 더욱 안정적인 출산율을 보여주는데 방대한 자료가 필요하고, 계산절차가 복잡하다는 단점이 있다.

본 연구에서는 연령별 합계출산율 추이를 시계열로 적합시켜 예측하기로 한다. 즉 모의 가임인구의 연도별 연령별 출산율 추이를 기초로 각세별 시계열분석을 실시한 후 예측한다. TFR 추이에서 나타나듯이 점진적으로 만혼화현상이 진행되고 있고, 출산율이 낮아지고 있다. 그리고 2000년도 들어서는 3아, 4아 이상의 출산순위는 매우 낮은 수준으로 머물러 있다.

시계열 분석을 위해 연령별 출산율 추이를 확률과정(stochastic process)으로 가정한다. 일반적으로 시계열은 정상성(stationary)을 보인다고 간주하지만 실제로 관측되는 대부분의 시계열자료는 정상성을 가지지 못하는 경우가 많다. 일반적으로 금융권 데이터나 인구 관련 데이터는 비정상시계열(non-stationary time series)이 대부분이므로 자료변환이나 차분을 통해서 정상시계열로 만든 후 분석한다.

본 연구에서는 정상·비정상을 판단하기 위해 Unit Test⁶⁾ 활용하고, 다음으로 차분여부를 결정하고 시계열 모형(ARMA, ARIMA)⁷⁾을 선택한다. 적합 및 진단과정을 거친 후 최적의 모형을 도출하여 예측을 하기로 한다. 이런 일련의 과정을 [그림 3-7]에 도시한다.



[그림 3-7] 시계열 적합 및 예측 과정

6) 시계열 정상성, 비정상성을 판단하는 검정을 말하며 상세한 설명은 <부록>을 참조
7) 시계열 모형에 대한 설명은 지면관계상 <부록>에 남기기로 한다.

먼저, TFR 시계열에 대한 Unit test 결과 15세~39세까지 출산율 추이는 비정상적으로 판정된다. 이를 정상으로 변환하기 위해 차분을 적용한 결과는 다음과 같다.

<표 3-2> TFR 차분 전·후의 정상비정상 판정 결과

Each	TFR	Each	TFR	Each	DIFF_TFR	Each	DIFF_TFR
age15	●	age31	●	age15	○	age31	○
age16	●	age32	●	age16	○	age32	○
age17	●	age33	●	age17	○	age33	○
age18	●	age34	●	age18	○	age34	○
age19	●	age35	●	age19	○	age35	○
age20	●	age36	●	age20	○	age36	○
age21	●	age37	●	age21	○	age37	○
age22	●	age38	●	age22	○	age38	○
age23	●	age39	●	age23	○	age39	○
age24	●	age40	○	age24	○	age40	○
age25	●	age41	○	age25	○	age41	○
age26	●	age42	○	age26	○	age42	○
age27	●	age43	○	age27	○	age43	○
age28	●	age44	○	age28	○	age44	○
age29	●	age45	○	age29	○	age45	○
age30	●	age46	○	age30	○	age46	○
		age47	○			age47	○
		age48	○			age48	○
		age49	○			age49	○
		stationary	○				
		nonstationary	●				

주, DIFF는 차분(differencing)을 뜻함.

정상·비정상 판정 후 모형수립을 위해서 ARMA(p,q), ARIMA(p,차수,q)를 결정한다. 모형의 식별단계는 p와 q를 결정하는 것이다. 이를 위해 모형선택의 기준(model selection criterion)으로 사용되는 통계량인 AIC(Akaike's Informaion Critetion) 또는 BIC(Bayesian Informaion Criterion) 등을 사용하는데 본 연구에서는 BIC를 사용한다.

모형의 식별과정을 통해 잠정적으로 모형이 선택되면 여러 모수추정법을 이용하여 모수를 추정해야한다. 널리 사용되는 적률추정법(method of moment estimation), 조건부 최소제곱추정법(conditional least squares estimation method), 비조건 최소제곱추정법(unconditional least squares estimation method), 최대가능도추정법(maximum likelihood estimation method) 등이 있다⁸⁾. 본 연구에는 최대가능도추정법을 이용한다.

만약 추정된 모수들이 모두 유의하면 다음 단계로 잠정모형이 과연 주어진 시계열 자료를 잘 설명하는가를 알아보게 되는데 이를 모형진단(model diagnostics)이라 한다. 모형진단 단계에서는 주로 잔차분석(residual analysis)과 과대적합진단(overfitting diagnostic)

8) 자세한 모수추정방법은 조신섭, 손영숙(2002), 「SAS/ETS를 이용한 시계열분석」의 9장을 참조



등의 분석⁹⁾을 한다. 이러한 분석의 결과가 이상 없을 경우에 최종예측모형으로 결정한다.

TFR, 출산순위별(1아, 2아, 3아 이상)에 따라 모형 적합, 진단의 결과를 살표보기 위해 시계열 분석을 한 결과를 <표 3-3>~<표 3-6>에 제시한다. 정상, 비정상을 위해 차분, 로그변화, 차분·로그변환 등을 적용하여 정상시계열로 만든 후, BIC 기준으로 모형을 선택하여, 포트멘토 통계량과 AIC통계량으로 잔차 및 과적합 진단 후 최적모형을 선택한다.

TFR 경우 가임연령의 저연령층(15~27세)는 1차 차분으로 정상시계열을 보였으며, 중간연령층(28~39세)은 1차 차분 후 로그변환으로 치환 후 정상시계열을 나타냈으며, 고연령층(40세 이상)은 차분, 치환없이 정상시계열이다.

출산순위별 1아인 경우 가임연령의 저연령층(15~27세)는 1차 차분으로 정상시계열을 보였으며, 중간연령층(28~38세)은 1차 차분 후 로그변환으로 치환 후 정상시계열을 나타냈으며, 39~47세까지는 1차 차분으로 정상시계열을 48~49세는 차분, 치환없이 정상시계열이다.

출산순위별 2아인 경우 1아의 경우에 거의 유사하며, 3아 이상인 경우 가임연령의 저연령층(15~26세)는 1차 차분으로 정상시계열을 보였으며, 중간연령층(28~34세)은 1차 차분 후 로그변환으로 치환 후 정상시계열을 나타냈으며, 35~49세까지 차분, 치환없이 정상시계열이다.

이러한 결과는 중간연령층의 출산을 추이가 확률적 추세를 가진다고 할 수 있으며, 어떤 경향(trend)을 보인다고 할 수 있다. 연령그룹별 출산력 추이를 살펴보면 29세 이전은 점진적 하강, 30~39세 집단은 점진적 상승으로 나타났다. 30대는 1986년 이후 U자형 출산율 추이를 보이고 있으며, 20대 후반에서 30대 초·중반의 출산연령이 이동하고 있다([그림 3-8] 참조). 이는 혼인연령의 상승 곧 만혼화의 진행을 고려하여 만혼, 만산의 연령별 출산율 패턴이라고 할 수 있다.

9) 자세한 모수추정방법은 조신섭, 손영숙(2002), 「SAS/ETS를 이용한 시계열분석」의 10장을 참조

〈표 3-3〉 TFR 시계열 변수변환, 차분, 적합, 최적모형 선정과정

연령	차분 및 변수변환 ¹⁾	model	BIC	잔차 및 과적합 진단 후 최적모형
age15	1차 차분	ARIMA(2,1,0)	-17.1605	ARIMA(2,1,0)
age16		ARIMA(0,1,3)	-16.5258	ARIMA(0,1,3)
age17		ARIMA(3,1,1)	-15.0356	ARIMA(3,1,1)
age18		ARIMA(1,1,3)	-13.6028	ARIMA(1,1,3)
age19		ARIMA(2,1,0)	-12.0691	ARIMA(2,1,0)
age20		ARIMA(1,1,3)	-10.8346	ARIMA(1,1,3)
age21		ARIMA(3,1,0)	-9.8853	ARIMA(3,1,0)
age22		ARIMA(3,1,2)	-9.9588	ARIMA(2,1,1), AIC(-239.3680)
age23		ARIMA(1,1,0)	-9.4285	ARIMA(1,1,0)
age24		ARIMA(3,1,0)	-8.8355	ARIMA(3,1,0)
age25		ARIMA(3,1,0)	-8.8646	ARIMA(3,1,0)
age26		ARIMA(1,1,0)	-8.8892	ARIMA(1,1,0)
age27		ARIMA(3,1,0)	-7.8301	ARIMA(3,1,0)
age28		1차 차분 1 로그변환	ARIMA(1,1,0)	10.2109
age29	ARIMA(1,1,0)		-4.8401	ARIMA(1,1,0), AIC(-78.5452)
age30	ARIMA(1,1,3)	-4.6562	ARIMA(0,1,2), AIC(-68.9891)	
age31	1차 차분	ARIMA(2,1,0)	-9.3224	ARIMA(2,1,0), AIC(-243.8730)
age32		ARIMA(3,1,0)	-9.7432	ARIMA(3,1,0), AIC(-259.9170)
age33		ARIMA(1,1,0)	-9.5953	ARIMA(1,1,1), AIC(-274.5890)
age34		ARIMA(1,1,1)	-10.2842	ARIMA(2,1,2), AIC(-293.6600)
age35	1차 차분 로그변환	ARIMA(3,1,0)	-3.8849	ARIMA(2,1,0), AIC(-41.6933)
age36		ARIMA(2,1,0)	-3.9001	ARIMA(2,1,0), AIC(-46.4958)
age37		ARIMA(2,1,0)	-3.8525	ARIMA(2,1,0), AIC(-43.9366)
age38	1차 차분	ARIMA(0,1,1)	-11.1375	ARIMA(1,1,2), AIC(-336.0250)
age39		ARIMA(1,1,0)	-11.7178	ARIMA(1,1,0)
age40	원자료	ARMA(1,3)	-11.9696	ARMA(1,3)
age41		ARMA(1,0)	-12.4346	ARMA(1,0)
age42		ARMA(3,1)	-12.6609	ARMA(3,1)
age43		ARMA(3,3)	-13.1027	ARMA(3,3)
age44		ARMA(2,0)	-13.2780	ARMA(2,0)
age45		ARMA(1,0)	-13.6584	ARMA(1,0)
age46		ARMA(2,0)	-14.0652	ARMA(2,0)
age47		ARMA(2,0)	-14.2618	ARMA(2,0)
age48		ARMA(2,0)	-14.6521	ARMA(2,0)
age49		ARMA(3,0)	-14.8270	ARMA(3,0)

주: 1) 차분은 윗줄에 변수변환은 아랫줄에 표시

〈표 3-4〉 10아 출산을 시계열 변수변환, 차분, 적합, 최적모형 선정과정

연령	차분 및 변수변환 ¹⁾	model	BIC	잔차 및 과적합 진단 후 최적모형
age15	1차 차분	ARIMA(2,1,0)	-17.5484	ARIMA(2,1,0)
age16		ARIMA(0,1,3)	-16.5652	ARIMA(0,1,3)
age17		ARIMA(3,1,1)	-15.1834	ARIMA(3,1,1)
age18		ARIMA(2,1,3)	-13.8353	ARIMA(2,1,3)
age19		ARIMA(1,1,0)	-12.6099	ARIMA(1,1,0)
age20		ARIMA(3,1,1)	-11.6025	ARIMA(0,1,1), AIC(-342.046)
age21		ARIMA(3,1,0)	-10.9983	ARIMA(3,1,0)
age22		ARIMA(3,1,3)	-11.0732	ARIMA(0,1,2), AIC(-289.890)
age23		ARIMA(0,1,0)	-10.6599	ARIMA(2,1,0), AIC(-277.333)
age24		ARIMA(3,1,0)	-10.0842	ARIMA(3,1,0)
age25		ARIMA(3,1,0)	-10.2755	ARIMA(3,1,0)
age26		ARIMA(0,1,0)	-9.9716	ARIMA(1,1,1), AIC(-297.654)
age27		ARIMA(0,1,0)	-9.8223	ARIMA(2,1,2), AIC(-294.835)
age28		1차 차분 1 로그변환	ARIMA(0,1,0)	8.2066
age29	1차 차분 로그변환	ARIMA(0,1,0)	-4.8811	ARIMA(2,1,2), AIC(-83.4971)
age30		ARIMA(0,1,0)	-4.5096	ARIMA(2,1,2), AIC(-60.9238)
age31	1차 차분	ARIMA(0,1,0)	-11.8738	ARIMA(1,1,1), AIC(-385.886)
age32		ARIMA(1,1,0)	-12.4380	ARIMA(1,1,0)
age33		ARIMA(0,1,0)	-12.7193	ARIMA(1,1,1), AIC(-427.085)
age34		ARIMA(1,1,1)	-13.3197	ARIMA(1,1,1)
age35	1차 차분 로그변환	ARIMA(0,1,0)	-3.8680	ARIMA(2,1,0), AIC(-39.5761)
age36	1차 차분	ARIMA(0,1,0)	-14.2084	ARIMA(2,1,0), AIC(-483.855)
age37	1차 차분 로그변환	ARIMA(0,1,0)	-3.9310	ARIMA(0,1,1), AIC(-39.983)
age38		ARIMA(0,1,0)	-15.0777	ARIMA(0,1,2), AIC(-25.0419)
age39	1차 차분	ARIMA(0,1,0)	-15.7318	ARIMA(1,1,1), AIC(-548.149)
age40		ARIMA(0,1,1)	-15.9710	ARIMA(0,1,1)
age41		ARIMA(0,1,0)	-16.6064	ARIMA(0,1,1), AIC(-588.412)
age42		ARIMA(0,1,1)	-17.6640	ARIMA(0,1,1)
age43		ARIMA(0,1,1)	-17.3384	ARIMA(0,1,1)
age44		ARIMA(0,1,1)	-17.6925	ARIMA(0,1,1)
age45		ARIMA(0,1,1)	-18.2704	ARIMA(0,1,1)
age46		ARIMA(0,1,1)	-18.5648	ARIMA(0,1,1)
age47		ARIMA(0,1,1)	-18.9700	ARIMA(0,1,1)
age48		원자료	ARMA(1,1)	-18.7509
age49	ARMA(3,0)		-19.3895	ARMA(3,0)

주: 1) 차분은 윗줄에 변수변환은 아랫줄에 표시

<표 3-5> 2아 출산을 시계열 변수변환, 차분, 적합, 최적모형 선정과정

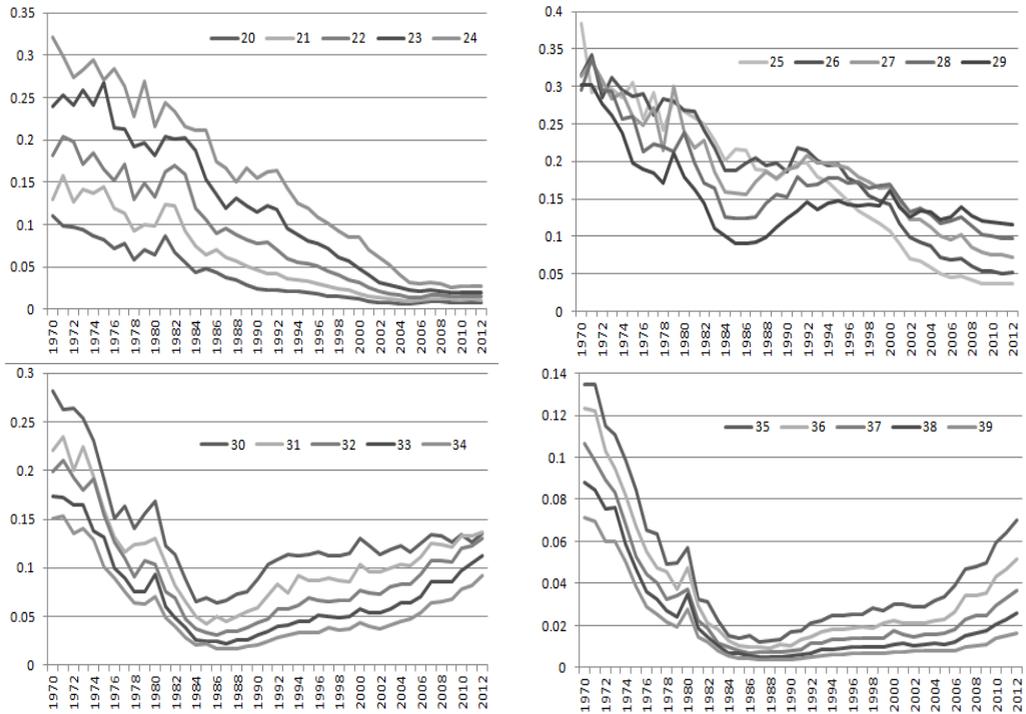
연령	차분 및 변수변환 ¹⁾	model	BIC	잔차 및 과적합 진단 후 최적모형
age15	1차 차분	ARIMA(0,1,0)	-21.0218	ARIMA(2,1,0), AIC(-771.201)
age16		ARIMA(0,1,3)	-20.2188	ARIMA(0,1,3)
age17		ARIMA(3,1,1)	-18.9325	ARIMA(3,1,1)
age18		ARIMA(0,1,3)	-16.9294	ARIMA(0,1,3)
age19		ARIMA(0,1,2)	-15.2048	ARIMA(0,1,2)
age20		ARIMA(3,1,0)	-13.1519	ARIMA(3,1,0)
age21		ARIMA(0,1,0)	-11.9169	ARIMA(2,1,0), AIC(-367.349)
age22		ARIMA(3,1,1)	-11.7768	ARIMA(2,1,2), AIC(-333.453)
age23		ARIMA(3,1,0)	-11.3781	ARIMA(3,1,0)
age24		ARIMA(0,1,3)	-10.4460	ARIMA(0,1,3)
age25		ARIMA(3,1,2)	-10.8382	ARIMA(3,1,2)
age26		ARIMA(0,1,0)	-10.5453	ARIMA(1,1,2), AIC(-306.151)
age27		ARIMA(1,1,0)	-9.3802	ARIMA(1,1,0)
age28		1차 차분 1 로그변환	ARIMA(1,1,0)	7.1516
age29	ARIMA(1,1,0)	6.0823	ARIMA(1,1,0)	
age30	1차 차분	ARIMA(0,1,0)	-10.505	ARIMA(3,1,0), AIC(-326.289)
age31	1차 차분 1 로그변환	ARIMA(0,1,0)	7.503	ARIMA(2,1,2), AIC(454.165)
age32	ARIMA(0,1,0)	10.475	ARIMA(2,1,2), AIC(576.625)	
age33	1차 차분	ARIMA(0,1,0)	-12.221	ARIMA(2,1,2), AIC(-392.401)
age34		ARIMA(0,1,1)	-12.643	ARIMA(2,1,0), AIC(-416.381)
age35	1차 차분 1 로그변환	ARIMA(0,1,0)	7.9337	ARIMA(2,1,2), AIC(478.357)
age36	1차 차분	ARIMA(1,1,1)	-13.899	ARIMA(1,1,1)
age37		ARIMA(1,1,1)	-14.510	ARIMA(1,1,1)
age38	1 1차 차분	ARIMA(0,1,0)	20.718	ARIMA(1,1,0), AIC(1020.369)
age39	1차 차분	ARIMA(2,1,0)	-15.832	ARIMA(2,1,0)
age40		ARIMA(0,1,0)	-16.330	ARIMA(1,1,2), AIC(-573.307)
age41		ARIMA(0,1,0)	-17.052	ARIMA(3,1,0), AIC(-597.061)
age42		ARIMA(3,1,0)	-17.484	ARIMA(3,1,0)
age43		ARIMA(1,1,0)	-17.714	ARIMA(1,1,0)
age44		ARIMA(0,1,1)	-18.144	ARIMA(0,1,1)
age45		ARIMA(1,1,0)	-18.833	ARIMA(1,1,0)
age46		ARIMA(0,1,1)	-18.674	ARIMA(0,1,1)
age47	원자료	ARMA(3,0)	-19.709	ARMA(3,0)
age48		ARMA(3,0)	-18.797	ARMA(3,0)
age49		ARMA(3,0)	-19.404	ARMA(3,0)

주: 1) 차분은 윗줄에 변수변환은 아랫줄에 표시

〈표 3-6〉 30아 이상 출산을 시계열 변수변환, 차분, 적합, 최적모형 선정과정

연령	차분 및 변수변환 ¹⁾	model	BIC	잔차 및 과적합 진단 후 최적모형
age15	1차 차분	ARIMA(3,1,0)	-23.875	ARIMA(3,1,0)
age16		ARIMA(0,1,1)	-20.584	ARIMA(0,1,1)
age17		ARIMA(3,1,0)	-22.884	ARIMA(3,1,0)
age18		ARIMA(2,1,0)	-22.257	ARIMA(2,1,0)
age19		ARIMA(3,1,1)	-20.315	ARIMA(3,1,1)
age20		ARIMA(0,1,3)	-16.289	ARIMA(0,1,3)
age21		ARIMA(0,1,3)	-15.017	ARIMA(0,1,3)
age22		ARIMA(0,1,3)	-13.758	ARIMA(0,1,3)
age23		ARIMA(0,1,1)	-13.336	ARIMA(0,1,1)
age24		ARIMA(3,1,3)	-12.419	ARIMA(2,1,0), AIC(-379.085)
age25		ARIMA(3,1,1)	-11.954	ARIMA(3,1,1)
age26		ARIMA(1,1,1)	-11.435	ARIMA(1,1,1)
age27		$\frac{1}{1차 차분}$	ARIMA(0,1,0)	16.359
age28	1차 차분	ARIMA(1,1,3)	-11.798	ARIMA(1,1,3)
age29	$\frac{1}{1차 차분}$	ARIMA(0,1,0)	16.018	ARIMA(2,1,2), AIC(814.681)
age30	원자료	ARMA(2,3)	-11.440	ARMA(2,3)
age31	1차 차분	ARIMA(3,1,3)	-11.719	ARIMA(0,1,2), AIC(-270.615)
age32		ARIMA(3,1,3)	-11.469	ARIMA(1,1,1), AIC(-293.547)
age33	$\frac{1}{1차 차분}$	ARIMA(1,1,0)	17.069	ARIMA(1,1,0)
age34		ARIMA(0,1,1)	16.441	ARIMA(0,1,1)
age35	원자료	ARMA(1,3)	-11.316	ARMA(1,3)
age36		ARMA(3,0)	-11.644	ARMA(3,0)
age37		ARMA(2,3)	-12.238	ARMA(2,3)
age38		ARMA(1,3)	-11.805	ARMA(1,3)
age39		ARMA(3,0)	-12.245	ARMA(3,0)
age40		ARMA(1,3)	-12.533	ARMA(1,3)
age41		ARMA(1,3)	-12.902	ARMA(1,3)
age42		ARMA(3,1)	-13.064	ARMA(3,1)
age43		ARMA(3,3)	-13.696	ARMA(3,3)
age44		ARMA(2,1)	-13.679	ARMA(2,1)
age45		ARMA(2,0)	-14.096	ARMA(2,0)
age46		ARMA(2,0)	-14.492	ARMA(2,0)
age47		ARMA(2,0)	-14.623	ARMA(2,0)
age48		ARMA(2,0)	-15.232	ARMA(2,0)
age49		ARMA(3,0)	-15.225	ARMA(3,0)

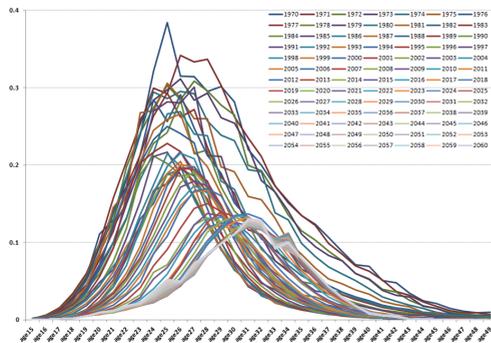
주: 1) 차분은 윗줄에 변수변환은 아랫줄에 표시



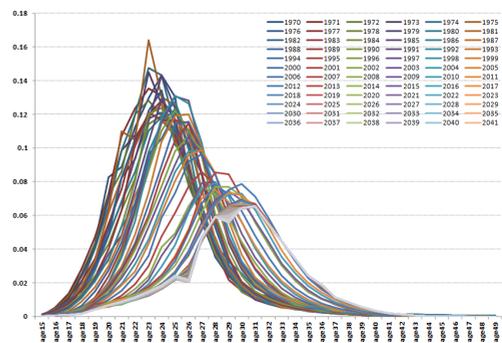
[그림 3-8] 연령별출산율(ASFR) 추이(1970~2012)

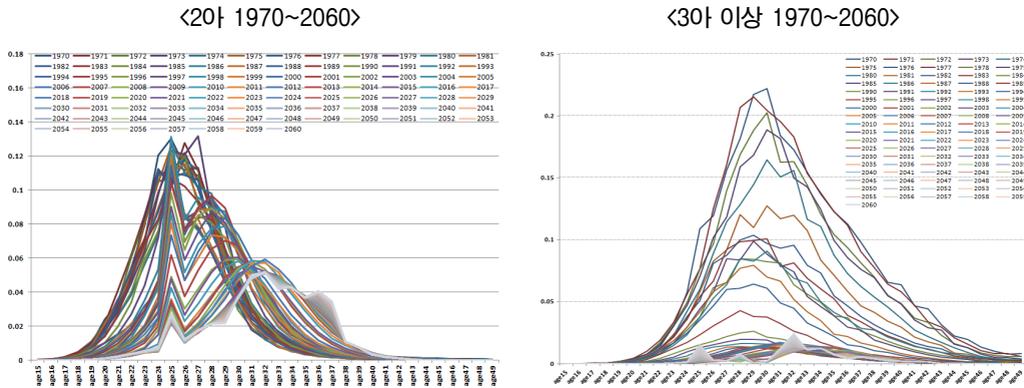
최적모형으로 선정된 시계열 결과를 활용하여 ASFR을 출산순위별로 도출하면 다음과 같다([그림 3-9] 참조). 모의 출산 연령의 점진적 상승과 분포가 넓게 퍼져 만혼화가 진행되고 있는 것을 살펴볼 수 있다.

<TFR 1970~2060>



<10아 1970~2060>





[그림 3-9] TFR 및 출산순위별 출산을 추이

2. 모수화모형

우리나라의 최근 장래인구추계에서는 합리적인 출생예측모형인 로그감마모형을 이용하여 출산순위별 완결출산율¹⁰⁾ 및 평균출산연령을 추정한다. 로그감마모형의 출발은 Stacy(1962)의 lifetime 분포모형¹¹⁾에서 시작한다. 그 뒤에 Prentice(1974), Farewell and Prentice(1977), Lawless(1980)가 분산 안정화 변화 및 delta-method 등의 수리적인 변화를 시도하여 확장된 로그감마모형(extended log gamma model)¹²⁾을 제안하였다.

<표 3-7> Lifetime 분포와 확장된 로그감마모형

<p>Lifetime 함수 (Stacy 1962)</p>	<p>Lifetime 분포 Stacy(1962) : $g(x) = \frac{\beta}{\Gamma(k)} \frac{x^{\beta k - 1}}{\alpha^{\beta k}} \exp\left[-\left(\frac{x}{\alpha}\right)^\beta\right], t > 0$</p>
<p>extended log gamma model (Prentice, 1974)</p>	<p>$g(w) = \frac{ q }{\Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp[q^{-2}\{(qw - \exp(qw))\}], -\infty < w < \infty,$ $w = \frac{x - \mu}{\sigma},$ 모수 q, μ, σ 3개</p>

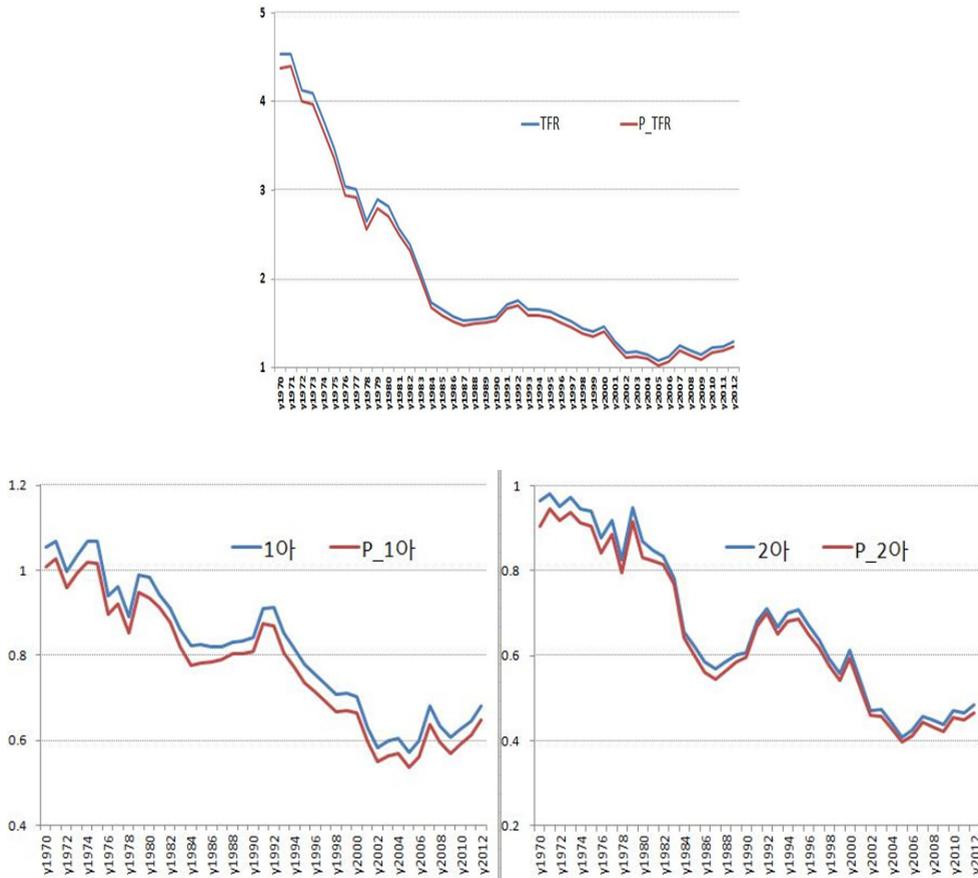
- 10) 합계출산율(한 여자가 가임기간 동안 낳을 것으로 예상되는 평균 출생수)과 비교되는 개념으로 출생 코호트별 완결출산율(completed fertility rate)는 특정연도에 태어난 여성 1명이 가임기간 동안 실제로 낳은 평균 출생아수를 가리킨다.
- 11) Stacy가 제안한 모형은 일반화 감마분포로서 생애분포에 적합하기 위해서 제안한 모형이다. 이 모형을 Prentice(1974)가 확장하여 extended log gamma model로 제안하였으며, 이 extended log gamma model을 일반화로그감마모형(GLG)이라고 한다.
- 12) 일반화 로그감마모형 도출관련 분포에 대해 <부록>에 제시함

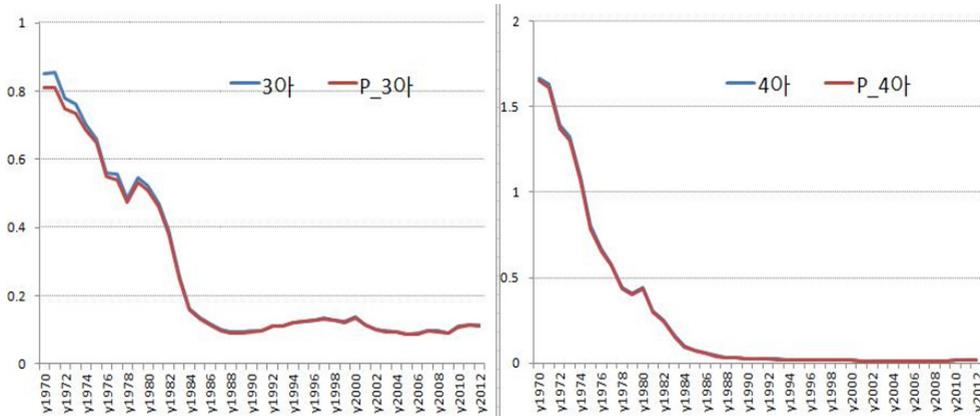
Kaneko(2003)는 출산순위별 C라는 확률값을 extended log gamma model에 적용하여 출산율모형을 제시하였는데 이를 일반화로그감마모형(Generalized Log Gamma Model; GLG)이라고 간주했다.

<표 3-8> 일반화로그감마모형

일반화 로그감마모형 (Generalized Log Gamma model)	출산력 모형에서의 GLG : $f(x) = Cg(x), \text{ transformation intergral } dw/dx = \frac{1}{\sigma},$ 모수4개, C는 출산순위의 확률
---	--

GLG모형을 활용하여 우리나라 출산순위별 적합결과는 [그림 3-10]와 같다.





주 : 'P'는 prediction을 의미함

[그림 3-10] 출산순위별 적합결과

모수화모형은 적합도가 낮고 모수 추정의 불안정과 같은 약점이 있어 출산율 예측에 적용하는데 약간의 단점이 있다. 모수화모형은 [그림 3-10]에서 살펴볼 수 있듯이 시점에 관계없이 실측치 TFR, 1아, 2아, 3아 등에서 과소추정(underestimate)하여 체계적 편향(systematic bias)을 보인다. 이러한 문제점인 과소추정을 어느 정도 보완하기 위해 새로운 출산율 추계방법이 필요하다.

이러한 단점을 보완하기 선진국의 출산장려정책을 참고하여 모델링(박유성, 2013)하거나, 출산율의 전이단계에서 출산율 감소율이 커질 때와 작아질 때가 대칭적이라는 강한 가정을 준다(Alkema et al, 2010).

그 외 미국, 유럽에서 출산율 예측을 위해 활용하는 모수화모형은 다음과 같다.

<표 3-9> 다양한 모수화모형

구분	모형식
Hadwiger함수 (Hadwiger 1940)	$f(x) = \frac{ab}{c} \left(\frac{c}{x}\right)^{3/2} \exp\left(-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right)$, 모수 a, b, c 3개
CM 함수 (Coale-McNeil, 1972)	$g(x) = \frac{\beta}{\Gamma(\alpha/\beta)} \exp[-\alpha(x-\mu) - \exp\{-\beta(x-\mu)\}]$, 모수 α, β, μ 3개, 초혼에 대한 연령분포를 적합한 모형 $f(x) = Cg(x) \leftarrow C$ is the proportion marrying, of a target cohort <i>first marriage schedule = C distⁿ of age at first marriage</i>



구분	모형식
베타함수 (Hoem et al. 1981)	$f(x) = R \frac{\Gamma(b+c)}{\Gamma(b)\Gamma(c)} (\beta-\alpha)^{(b+c-1)} (x-\alpha)^{b-1} (\beta-x)^{(c-1)},$ 모수 R, b, c 3개
혼합 Hadwiger (Cahandola et al, 1999)	$f(x) = m \left(\frac{b_1}{c_1}\right) \left(\frac{c_1}{x}\right)^{3/2} \exp\left(-b_1^2 \left(\frac{c_1}{x} + \frac{x}{c_1} - 2\right)\right) + (1-m) \left(\frac{b_2}{c_2}\right) \left(\frac{c_2}{x}\right)^{3/2} \exp\left(-b_2^2 \left(\frac{c_2}{x} + \frac{x}{c_2} - 2\right)\right)$ 모수 m, b_1, b_2, c_1, c_2 5개
PK1 함수(Periatera and Kostaki 2007)	$f(x) = c \exp\left(-\left(\frac{x-\mu}{\sigma(x)}\right)^2\right), \begin{cases} \sigma(x) = \sigma_1, & \text{if } x \leq \mu \\ \sigma(x) = \sigma_2, & \text{if } x > \mu \end{cases}, \quad c = \sum_{i=15}^{49} ASFR_i,$ 모수 $c, \mu, \sigma_1, \sigma_2$ 4개,
PK2 함수(Periatera and Kostaki 2007)	$f(x) = c_1 \exp\left(-\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right) + c_2 \exp\left(-\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right),$ $c_1 = \sum_{i=15}^{\mu} ASFR_i, c_2 = \sum_{\mu}^{49} ASFR_i$ 모수 $c_1, c_2, \mu_1, \mu_2, \sigma_1, \sigma_2$ 6개, 쌍봉 정상(bimodal peaks)을 가진 출생률 자료에 적합

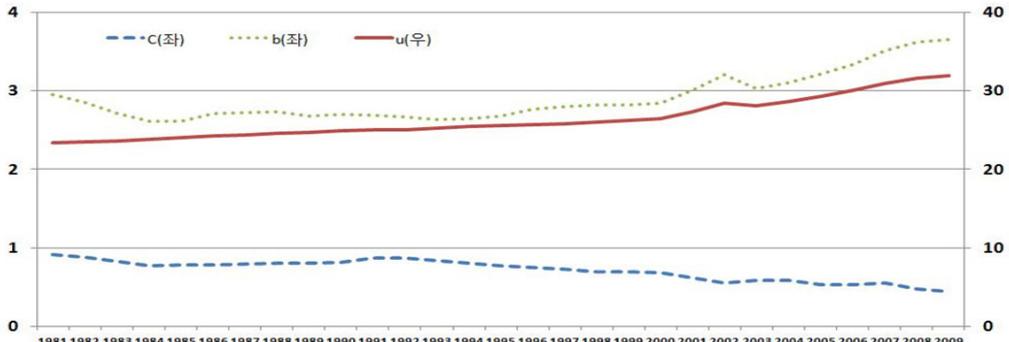
혼합 Hadwiger 및 PK2 함수는 10대 후반에도 출생률이 높게 나타나는 쌍봉 정상(bimodal peaks)을 가진 출생률 자료를 적합하기 위해 제안된 모수화모형이다. 우리나라 출산율은 쌍봉 정상형태가 아니므로 PK2, 혼합 Hadwiger 함수를 이용하여 신뢰성 있는 출산율의 예측을 기대할 수 없음을 의미한다. CM 함수는 초혼(Nuptiality)에 대한 연령별 자료를 적합하기 위한 모수화모형이며, Gamma, GLG, Hadwiger, 베타함수는 연령별 산모의 평균연령과 표준편차를 구할 수 있는 모형(Hoem et al., 1981)이다.

모수화모형 변화가 전형적인 ASFR의 패턴을 반영하고 있다. 출산율이 가장 높은 연령을 중심으로 좌우대칭의 형태였으나 이런 형태는 점차적으로 고연령의 출산율이 저연령대 출산율보다 높아지는 형태(독일, 프랑스, 일본, 한국 등) 또는 10대 후반의 출산율이 예외적으로 높아지는 쌍봉형태(bimodal peaks, 영국 및 미국)로 변형되고 있다.

또한 이러한 현상을 위에 제시된 모수화모형의 적합에서도 확인할 수 있는데, 모수 추정 시 평균 출산연령과 표준편차는 증가추세를 따르고 있다. 모수화모형만으로 출산율을 예측하면 평균 출산연령은 증가하고 분산은 커지게 도출되어 미래의 출산율은 감소할

수밖에 없는 결과가 나온다.

[그림 3-11]에서 볼 수 있듯이 우리나라 산모의 평균연령과 표준편차는 증가추세를 보이고 있다.



주 : 표준편차가 점진적으로 증가하는 의미는 가임연령이 점점 넓게 퍼지면서 동시에 고연령대 임신이 늘어나는 것을 뜻함.

[그림 3-11] GLG모형 모수추정[출산순위 C, 모의 평균출산연령(μ), 표준편차(b)]

이상으로 논의된 출산력 모형은 다섯 가지 특징으로 요약이 가능하다.

- ① 출산율이 가장 높은 모연령(μ)을 중심으로 대칭 또는 우측출산율 증가형태
- ② 모연령(μ)을 중심으로 평균(location)은 동일하나 분포(dispersion)와 스케일(scale)이 다름
- ③ 스케일은 모연령을 중심으로 좌(c₁), 우(c₂)의 합계출산율을 활용
- ④ 분포는 각 코호트의 평균출산연령의 표준편차로 접근
- ⑤ PFR, CFR 모두 지수형태(exponential type)를 보이므로 exp함수 포함

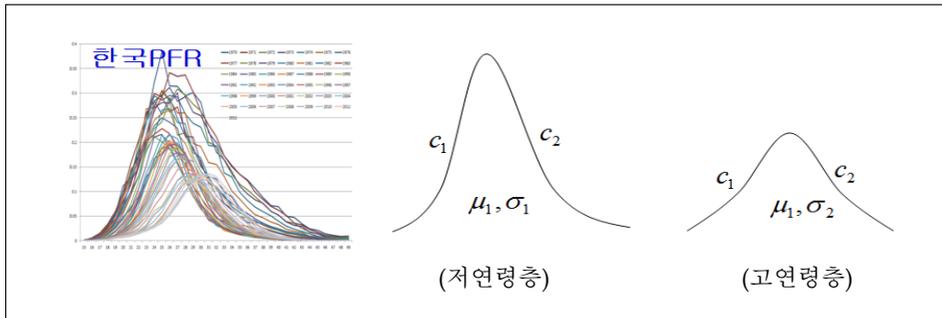
이러한 특징을 잘 반영하기 위해 모의 평균출산연령을 중심으로 서로 다른 두 개의 분포가 합쳐진 형태인 아래의 모형식을 제안한다.

제안 모형식 $f(x) = c_1(1) + c_2(2) + \epsilon_i, i = 1970, \dots, 2013$

주 : (1), (2)에 지시함수와 모평균(μ)을 중심으로 좌·우 모형이 필요
 이를 위해 지시변수(indicator variable, $I_{(x < \mu)}$, $I_{(x \geq \mu)}$)를 적용



c_1 은 당해연도 모의 가장 높은 출산연령(μ)보다 작은 연령에서의 합계출산율이고, c_2 는 μ 보다 큰 연령에서의 합계출산율이다. [그림 3-12]에서 볼 수 있듯이 우리나라는 μ 를 중심으로 좌우형 대칭인 단일봉형태를 유지하고 있다.



[그림 3-12] 저연령층, 고연령층의 기간합계출산율(Period-TFR)

다음으로 ‘제안 모형식’의 (1)과 (2)에 적합한 분포를 선택하는 문제가 남아 있는데, 본 연구에서는 Hybrid PK1·PK2, Folded-normal, Normal, GLG(C 제외), Cauchy 분포를 제안한다.

이유는 앞에서 제시한 5가지 특징을 고루 갖춘 분포이며 위치모수(location parameter, 대표적인 mean)와 척도모수(scale parameter, 대표적인 variance)의 적용이 용이하기 때문이다. 위치모수를 중심으로 척도모수를 조정하여 대칭 비대칭 분포를 만들 수 있다. 보다 자세한 근거는 다음과 같다.

첫째, Hybrid PK1·PK2는 PK1모형과 PK2 모형을 혼용한 것으로 모의 높은 출산연령 좌우의 c_1, c_2 가 적용가능하며 ‘exp’의 지수함수를 포함하고 있다.

둘째, Folded Normal 분포는 normal 분포와 반 접은 형태로 좌우분포를 적합할때 매우 유용한 분포이며, c_1, c_2 가 적용가능하고 ‘exp’의 지수함수를 포함하고 있다.

셋째, GLG(C 제외)모형은 부정확한 출산순위 확률값 C를 제외한 후, 그 대신에 모의 출산연령 좌우의 c_1, c_2 를 대입하여 출산율 예측을 한다.

넷째, 우리나라가 만혼화 현상이 진행되고 있고, 모수화모형의 약점인 과소추정을 보완하기 위해서 정규분포보다 꼬리가 두터운 Cauchy 분포를 제안한다.

〈표 3-10〉 출산율 예측 제안 모수화모형

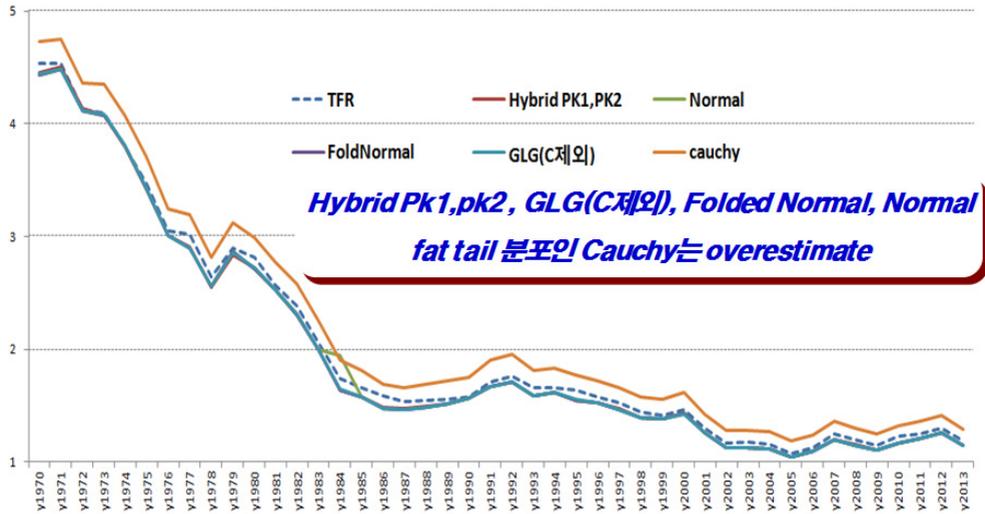
제안 모형 (모수 4개, $c_1, c_2, \sigma_1 = b_1, \sigma_2 = b_2$)	
GLG distribution (C제외)	$f(x; c_1, c_2, b_1, b_2) = c_1 \frac{ \lambda }{b_1 \Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left[\frac{1}{\lambda} \left(\frac{x-\mu}{b_1} \right) - \frac{1}{\lambda^2} \exp \left(\lambda \left(\frac{x-\mu}{b_1} \right) \right) \right] I_{[x \leq \mu]}$ $+ c_2 \frac{ \lambda }{b_2 \Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp \left[\frac{1}{\lambda} \left(\frac{x-\mu}{b_2} \right) - \frac{1}{\lambda^2} \exp \left(\lambda \left(\frac{x-\mu}{b_2} \right) \right) \right] I_{[x > \mu]}$
Folded Normal distribution	$f(x; c_1, c_2, \sigma_1, \sigma_2) = c_1 \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left(-\frac{(-x-\mu)^2}{2\sigma_1^2} \right) I_{[x \leq \mu]} + c_2 \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left(-\frac{(x-\mu)^2}{2\sigma_2^2} \right) I_{[x > \mu]}$

주: Normal 분포는 Folded Normal 분포에서 좌측부분($-x$)의 부호를 '+' (x)로 변경하면 된다.

제안 모형 (모수 4개, $c_1, c_2, \sigma_1 = b_1, \sigma_2 = b_2$)	
Hybrid PK1, PK2	$f(x; c_1, c_2, \sigma_1, \sigma_2) = c_1 \exp \left(-\frac{(x-\mu)^2}{\sigma_1^2} \right) I_{[x \leq \mu]} + c_2 \exp \left(-\frac{(x-\mu)^2}{\sigma_2^2} \right) I_{[x > \mu]}$
Cauchy distribution	$f(x; c_1, c_2, \sigma_1, \sigma_2) = c_1 \frac{\sigma_1}{\pi} \frac{1}{\sigma_1^2 + (x-\mu)^2} I_{[x \leq \mu]} + c_2 \frac{\sigma_2}{\pi} \frac{1}{\sigma_2^2 + (x-\mu)^2} I_{[x > \mu]}$

다음으로 1970~2013년 TFR을 대입하여 제안한 모수화모형 적합도를 알아보기 위해 실측치와 어느 정도 차이가 있는지 확인한다. 적합결과 Cauchy> TFR(실측치)> Hybrid PK1·PK2> Folded-Normal Normal> GLG(C제외) 순으로 나타난다. 하지만 2000년도 이후의 적합치와 실측치는 거의 동일선상에 놓여 있다는 것을 확인할 수 있다(〈표 3-11〉참조).

그리고 다섯 모수화모형의 적합치들의 평균은 1970~2013년도 실측치와 거의 일치하며, 적합정도는 좋다고 할 수 있다.



[그림 3-13] 제안 모수화모형 적합결과 추이

<표 3-11> 제안 모수화모형 적합결과와 TFR과 평균

	y2000	y2001	y2002	y2003	y2004	y2005	y2006
TFR	1.4669	1.2974	1.1662	1.1801	1.1540	1.0764	1.1231
Hybrid PK1,PK2	1.4313	1.2602	1.1246	1.1277	1.1204	1.0410	1.0987
Normal	1.4296	1.2589	1.1243	1.1275	1.1197	1.0405	1.0986
FoldNormal	1.4296	1.2589	1.1243	1.1275	1.1197	1.0405	1.0986
GLG(C제외)	1.4278	1.2582	1.1245	1.1282	1.1179	1.0392	1.0949
cauchy	1.6162	1.4253	1.2745	1.2838	1.2722	1.1848	1.2434
mean	1.4669	1.2923	1.1544	1.1589	1.1500	1.0692	1.1268

	y2007	y2008	y2009	y2010	y2011	y2012	y2013
TFR	1.2501	1.1924	1.1489	1.2255	1.2439	1.2967	1.1868
Hybrid PK1,PK2	1.1958	1.1516	1.1041	1.1712	1.2093	1.2599	1.1477
Normal	1.1954	1.1506	1.1033	1.1707	1.2085	1.2591	1.1473
FoldNormal	1.1954	1.1506	1.1033	1.1707	1.2085	1.2591	1.1473
GLG(C제외)	1.1945	1.1477	1.1010	1.1693	1.2049	1.2560	1.1450
cauchy	1.3604	1.3027	1.2509	1.3223	1.3565	1.4152	1.2873
mean	1.2283	1.1807	1.1325	1.2008	1.2375	1.2898	1.1749

주 : 'mean'은 제안한 모수화모형의 적합지들의 평균을 의미



3. 출산율 예측(2014~2060년)

일반적으로 출산율 예측을 위해 코호트별 일반화 로그감마분포모형에 적용할 출생순위별 평균출산연령의 변화를 저위, 저위, 고위 등의 3가지로 가정한다.

모수화모형으로 출산율 예측 시 저위, 중위, 고위 가정, 각 모수추정(GLG의 경우; C, μ, σ, λ), 시계열 예측 등 많은 계산량을 필요로 한다. 또한 완결출산율을 생성하기 위해 미완결 출생 코호트부터는 더 이상 혼인 연령이 상승하지 않는다는 가정과 코호트의 혼인 행태나 출산행태에 대하여 확실한 정보가 부재하다. 이러한 가정은 출생순위별 평균출산연령과 모든 자녀에 대한 평균출산연령은 해당연도의 혼인연령이 상승하는 만큼의 효과가 있다는 사실을 명심할 필요가 있다(전광희, 2006).

이러한 혼인연령의 상승과 독신경향의 증가로 인한 평균출산연령의 증가는 중위수준과 저위수준에는 그대로 적용된다. 그러나 고위수준에 대해서는 최근의 혼인연령 상승이 상대적으로 일찍부터 완만해진다고 보고, 어느 특정한 출생코호트에 대하여 매년 어느 정도의 양으로 상승한다고 가정한다.¹³⁾

모수화모형만으로 예측하는 방법의 단점으로 적합도와 모수 추정의 불안정성, 과거 출산율 패턴이 미래에도 지속되어야 한다는 가정을 언급하였다. 그리고 출생순위마다 독립적으로 파라미터를 추정하여 계산량도 많아지고 정확성도 보장하기 힘들다.

특히 과거 출산율 패턴이 미래에도 지속되어야 한다는 가정이 우리나라와 같이 최근 까지 출산율이 감소추세인 경우에는 출산율이 감소하여 정부에서 시행하고 있는 선택적 복지를 반영하기가 힘들다.

따라서 본 연구에서는 이러한 가정에서 벗어나, 우리나라 경우 만혼화가 진행되고 있고 30대 초·중반 높은 출산율을 반영할 수 있는 ASFR 개별 출산율 추이를 반영한 시계열 예측치 결과를 모수화모형에 대입시켜 최종적인 출산순위별 출산율을 예측해 본다(그림 3-14 참조).



[그림 3-14] 출산순위별 예측값 도출 과정

13) 보다 자세한 설명은 전광희(2006), “장래인구추계를 위한 출산력 가정치의 설정”의 pp.65~66 참고

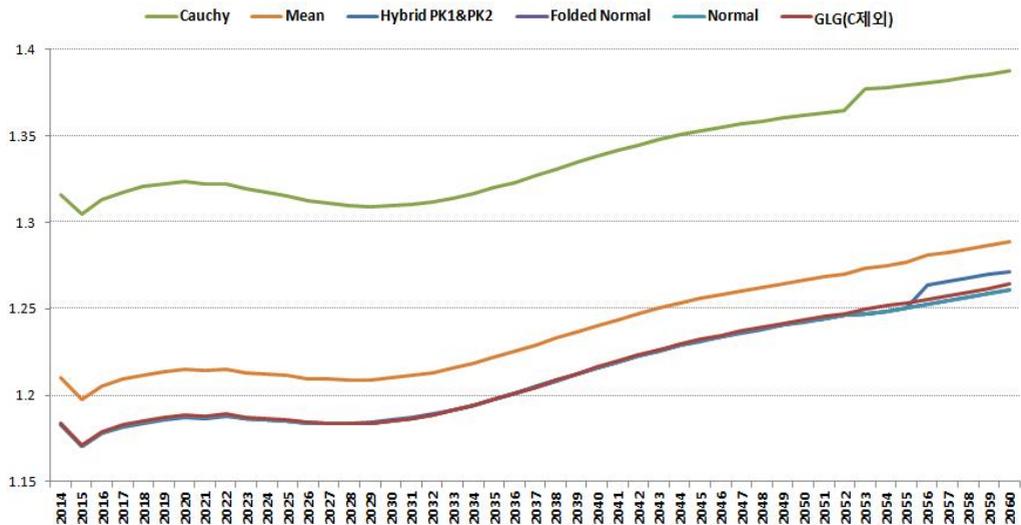
TFR, 1아, 2아, 3아 이상¹⁴⁾의 연령별 시계열 예측치가 제안된 모수화모형에 대입되어 2014년 이후의 예측치를 도출한 결과는 다음과 같다.

시계열 예측결과를 반영한 모수화모형 출산율 예측을 종합적으로 살펴보면, 2014년에는 합계출산율이 여자 1인당 1.183~1.316명, 2040년 1.216~1.338명, 2060년 1.261~1.388명에 이르게 된다.

통계청의 2010년 장래인구추계와 비교한 결과 본 연구결과의 향후 출산율은 현수준과 중위 사이에 위치한 것으로 예측된다. 저위, 중위는 혼인연령의 상승 및 독신경향의 증가로 이는 평균출산연령의 증가를 의미한다.

<표 3-12> 다양한 모수화모형의 TFR 예측 결과와 장래인구추계 비교

다양한 모수화모형 TFR 예측 결과							장래인구추계			
연도 모형	2014	2020	2030	2040	2050	2060	가정	2010	2040	2060
Cauchy	1.316	1.324	1.309	1.338	1.362	1.388	고위	1.23	1.79	1.79
평균	1.210	1.215	1.210	1.240	1.266	1.289				
Hybrid PK1·PK2	1.183	1.187	1.185	1.216	1.242	1.272	중위	1.23	1.42	1.42
Folded-Normal	1.183	1.188	1.185	1.216	1.243	1.261				
Normal	1.183	1.188	1.185	1.216	1.243	1.261	저위	1.23	1.01	10.1
GLG(제외)	1.184	1.215	1.185	1.216	1.243	1.264	현수준	1.23	1.19	1.19



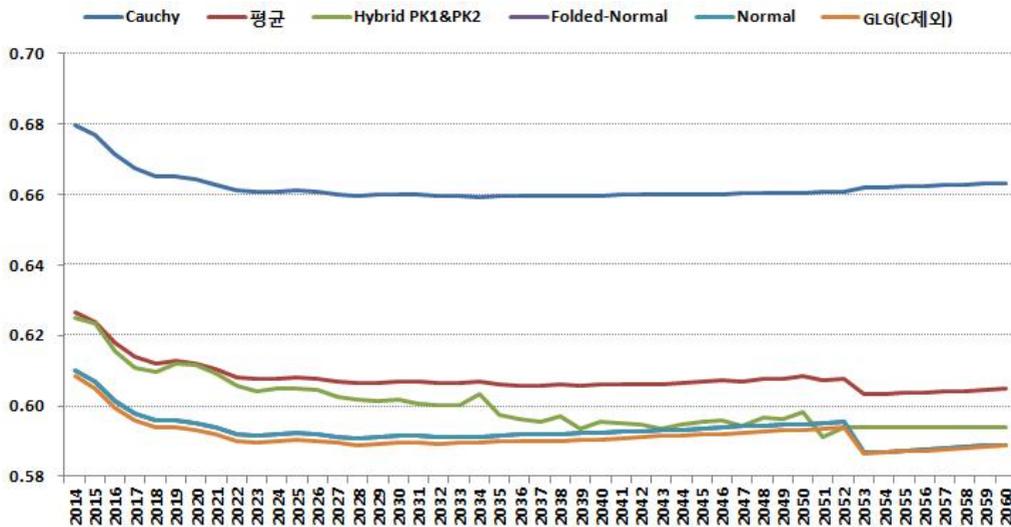
[그림 3-15] 2014~2060년 합계출산율 추이

14) 3아, 4아 이상의 출산율이 합계출산율에서 차지하는 비중과 변화가 미비하므로 3아 이상으로 분석

1아의 출산율 예측 결과 예측 시작점인 2014년에는 여자 1인당 0.608~0.680명을 보이다가 약간씩 감소하여 2040년 0.591~0.660명, 2060년 0.589~0.663명에 이르게 된다.

〈표 3-13〉 다양한 모수화모형의 1아 예측 결과

다양한 모수화모형 TFR 예측 결과						
모형 \ 연도	2014	2020	2030	2040	2050	2060
Cauchy	0.680	0.664	0.660	0.660	0.661	0.663
평균	0.627	0.612	0.607	0.606	0.608	0.605
Hybrid PK1-PK2	0.625	0.612	0.602	0.596	0.598	0.594
Folded-Normal	0.610	0.595	0.591	0.592	0.595	0.589
Normal	0.610	0.595	0.591	0.592	0.595	0.589
GLG(C제외)	0.608	0.593	0.589	0.591	0.593	0.589

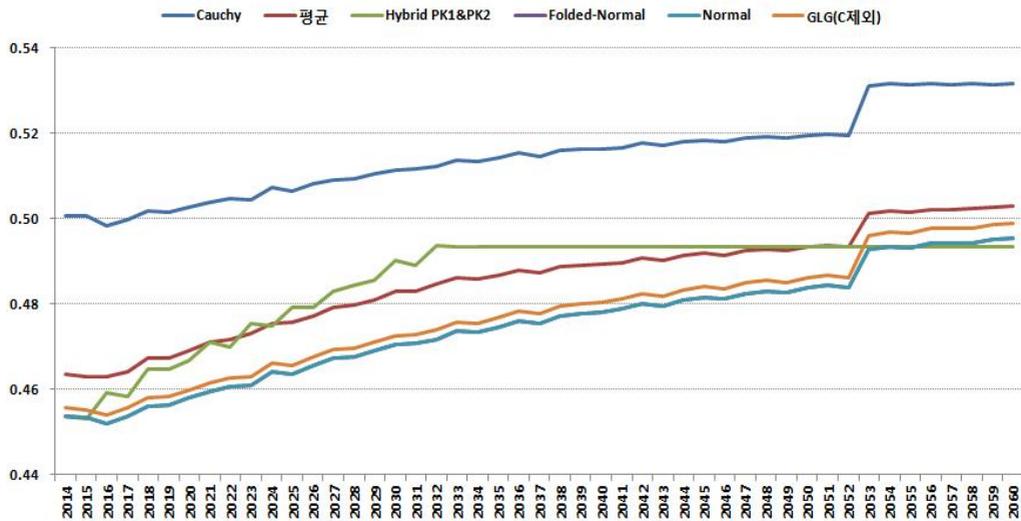


[그림 3-16] 2014~2060년 출산순위 1아 출산율 추이

2아의 출산율 예측 결과 예측 시작점인 2014년에는 여자 1인당 0.454~0.501명을 보이다가 약간씩 상승해 2040년 0.480~0.516명, 2060년 0.493~0.532명에 이르게 된다.

〈표 3-14〉 다양한 모수화모형의 2아 예측 결과

다양한 모수화모형 TFR 예측 결과						
모형 \ 연도	2014	2020	2030	2040	2050	2060
Cauchy	0.501	0.503	0.512	0.516	0.520	0.532
평균	0.463	0.469	0.483	0.489	0.493	0.503
Hybrid PK1·PK2	0.454	0.467	0.490	0.493	0.493	0.493
Folded-Normal	0.454	0.458	0.471	0.478	0.484	0.495
Normal	0.454	0.458	0.471	0.478	0.484	0.495
GLG(C제외)	0.456	0.460	0.473	0.480	0.486	0.499

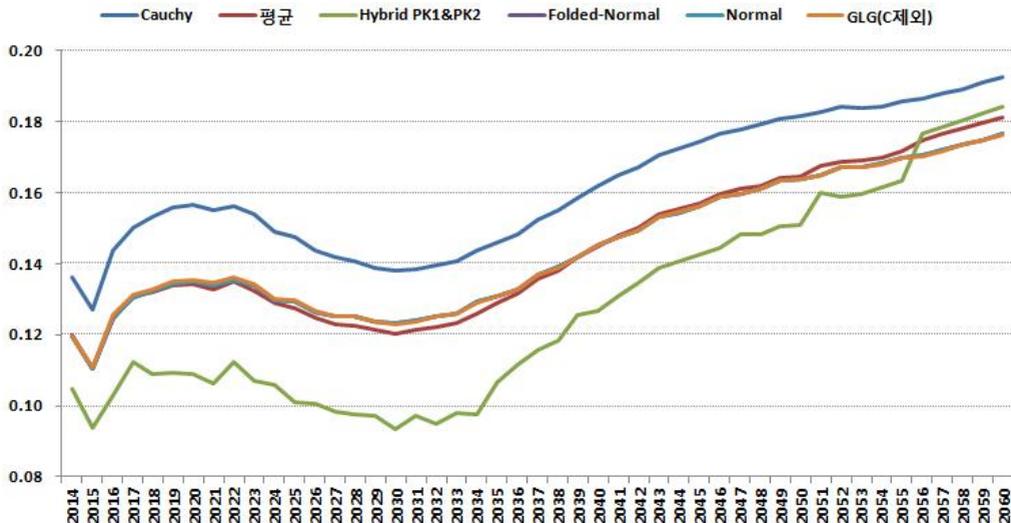


[그림 3-17] 2014~2060년 출산순위 2아 출산율 추이

3이 이상의 경우는 모수화모형의 결과는 너무 변동이 심하고 현실적인 측면(‘3아 이상 출산율= TFR-합계(1아, 2아 출산율)’ 기준에는 합당하지 않은 결과가 도출되어 약간의 보정작업을 거쳐 제시한다. ASFR 모형 수정과 기준에 입각해서 재추정한 결과 예측 시작점인 2014년에는 여자 1인당 0.105~0.136명을 보이다가 약간씩 상승해 2040년 0.127~0.162명, 2060년 0.176~0.193명에 이르게 된다.

〈표 3-15〉 다양한 모수화모형의 3아 이상 예측 결과

		다양한 모수화모형 TFR 예측 결과					
모형	연도	2014	2020	2030	2040	2050	2060
Cauchy		0.136	0.156	0.138	0.162	0.182	0.193
평균		0.120	0.156	0.138	0.162	0.182	0.193
Hybrid PK1·PK2		0.105	0.109	0.103	0.127	0.151	0.184
Folded-Normal		0.119	0.135	0.123	0.145	0.164	0.177
Normal		0.119	0.135	0.123	0.145	0.164	0.177
GLG(C제외)		0.120	0.136	0.123	0.145	0.164	0.176



[그림 3-18] 2014~2060년 출산순위 3아 이상 출산율 추이

시계열 예측결과를 반영한 모수화모형 출산율 예측을 종합적으로 살펴보면, Cauchy 분포에 의한 출산율이 과대추정(overestimate)되는 경향이 있는 반면 그 외 분포는 전반적으로 과소추정(underestimate)의 결과가 도출된다.

과대, 과소추정의 보안을 위한 평균(모수화모형 예측치들의 평균)치를 살펴보면 출산율 예측 시작점인 2014년에는 합계출산율이 여자 1인당 1.210명으로, 2040년 1.240명, 2060년 1.289명에 이르게 된다.

1아는 2014년 여자 1인당 0.627명을 보이다가 약간 주춤하여 2040년 0.606명, 2060년

0.605명에 이르게 된다. 2아는 2014년 여자 1인당 0.463명을 보이다가 약간씩 상승해 2040년 0.489명, 2060년 0.503명에 이르게 된다. 3아 이상은 2014년 여자 1인당 0.120명을 보이다가 약간 상승해 2040년 0.162명, 2060년 0.193에 이른다.

그리고 1970~2013년 출산율 실측치에 대한 모수화모형 적합결과 적합치(Hybrid PK1·PK2 적합치, Folded Normal 적합치, Normal 적합치, GLG(C제외) 적합치, Cauchy 적합치)의 평균과 거의 일치한다는 것을 참고한다면, 향후 출산율 예측은 다양한 모수화 모형의 예측치 평균을 제시하는 것도 하나의 방안이 될 수 있다고 본다.

제3절 결론 및 시사점

본 연구에서는 출산율 예측에 대해서 시계열분석, 모수화모형을 활용하여 2014~2060년까지 합계출산율, 1아, 2아, 3아 이상을 도출하였다.

모수화모형만으로 추계하는 경우의 단점을 ASFR 추이를 반영한 시계열 예측값을 모수화모형에 대입하여 보완하였다. Hybrid PK1·PK2, Folded-normal, Normal, GLG(C제외), Cauchy분포를 기초한 모수화모형에 출산율 추이를 내포하고 있는 시계열 분석결과를 입력치로 간주하여 향후 출산율을 예측하였다. 또한 이 결과를 2010년 통계청의 장래인구추계 출산율과 비교한 결과 현수준과 중위수준에 있는 것으로 나타났다. 이는 평균출산연령 상승과 독신여성의 증가를 고려한 장래인구추계 저위·중위 가정을 반영한 결과라 볼 수 있다.

선행 연구들과 다르게 1아 출산율이 약간 감소되는 경향이 있고 3아 이상의 출산율이 상대적으로 높은 값으로 도출되었다. 하지만 3아 이상의 출산율은 0.2 미만으로 미비하므로 전체 합계출산율 예측에 큰 영향을 미치지 못한다. 문제는 1아, 2아인데 1아는 다소 감소, 2아는 증가이나, 향후 미래에 대한 출산율 예측이 본 결과와 같이 전개될 것이라고는 장담할 수는 없다. 본 결과는 시계열 분석과 모수화모형에 입각한 예측결과이므로 향후 예측력을 위한 정밀한 작업이 필요하다고 본다.

끝으로 향후 연구분야를 제시해 본다면, 시나리오(저위, 중위, 고위) 추계에서 현실적인 정책 효과를 반영한 추계모형이 필요하다고 본다. 확률적 구간인 저(중, 고)위를 제시하는 것도 중요하지만, 출산장려정책 또한 이벤트를 반영할 수 있는 모형과 출산율의 변화와 향후 목표치를 보여줄 수 있는 시뮬레이터(simulator)가 요구된다.

선진국의 출산장려정책을 참고하여 우리나라의 향후 출산율 목표치 및 벤치마킹 등의 참고자료로 활용하는 것도 유용한 연구라 생각된다.



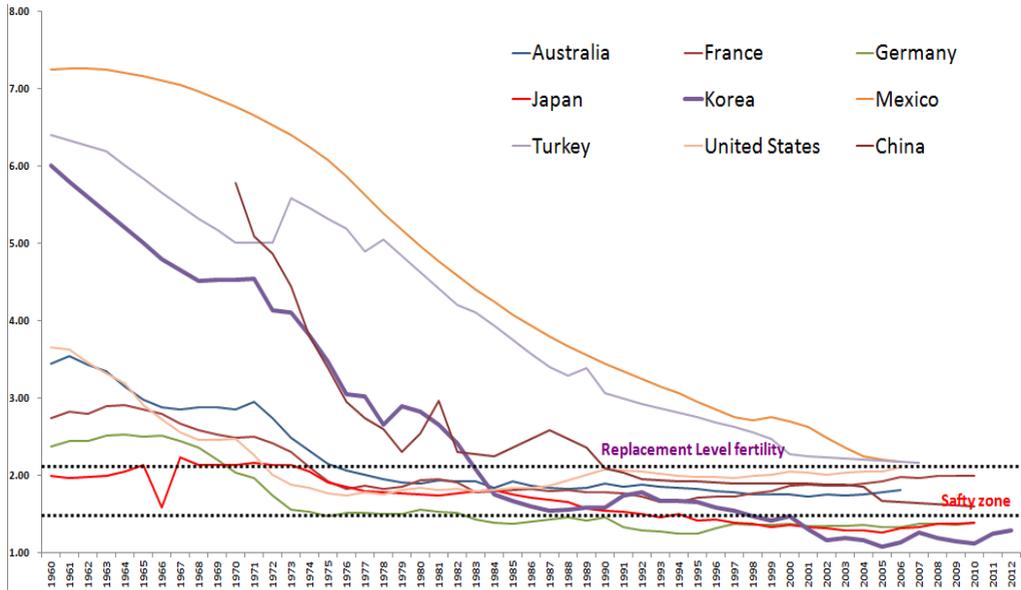
<참고문헌>

- 박유성, 김미리, 김성용(2013), “확률적 출산율 모형과 한국의 미래인구 구조”, 조사연구, 14권 3호, pp.49~78.
- 이민숙(2014), “프랑스 저출산 문제 해소 요인”, 「한국인구학회 후기 학술대회」, pp.82~91.
- 이상열(2013), 「시계열 분석, 이론 및 SAS 실습」, 자유아카데미.
- 전광희(2003), “한국의 출산력 변천:추이와 예측”, 통계연구, 제8권 제1호, pp.33-58
- 전광희(2006), “장래인구추계를 위한 출산력 가정치의 설정”, 한국인구학 제29권 제2호, pp.53~88.
- 조신섭, 손영숙(2002), 「SAS/ETS를 이용한 시계열 분석」, 율곡출판사.
- 통계교육원(2013), 「인구통계 기초와 응용」.
- 통계청(2011), 「2011년 인구동태통계연보 (총괄·출생·사망편)」
- 통계청(2012), 「2012년 인구동태통계연보 (총괄·출생·사망편)」
- Alkema, L., et al.(2011), "Probabilistic Projections of the Total Fertility Rate for All Countries." Demography 48, pp.815~839.
- Coale, A. and D. R, McNeil(1972), "The Distribution by age of the Frequency of First Marriage in Female Cohort." journal of American Statistical Association 36, pp.743~749.
- Farewell, V. T. and Prentice, R. L.(1977), "A study of distributional shape in life testing", Technometrics, 19, pp.69~75.
- Hadwiger, H.(1940), "Eine Analytische Reproductions-Funktion fur Biologische Gesamtheiten." Skandinavisk Aktuarietidskrift 23: pp.101~113.
- Hoem et al(1981), "Experiments in Modeling Recent Danish Fertility Curves." Demography 18, pp.231~244.
- Kaneko, Ruichi(2003), "Elaboration of the Coale-McNeil Nuptiality Model as the Generalized Log Gamma Distribution: A New Identity and Empirical Enhancements," Demographic Research 9, Article 10.
- Lawless, J.F.(1980), "Inference in the Generalized Gamma and Log Gamma Distributions", Technometrics, VOL.22, No.3.
- Lutz, W., V. Skirbekk, M. Testa.(2006), "The Low Fertility Trap Hypothesis: Forces that May Lead to Further postponement and Fewer Births in Europe." Vienna Yearbook of population Research 2006.
- MaDonal, P.(2006). "Low Fertility and the State: The Efficacy of Policy." Population and Devopment Review 32(3): 401~510.
- McDonald(2007), "Low Fertility and Policy", AGEING HORIZONS, issue No.7, pp.22~27.4

- Prentice, R. L.(1974), "A log-gamma model and its maximum likelihood estimation", *Biometrika*, 61, pp.539~544.
- Peristera & Kostaki(2007), "Modeling fertility in modern populations", *DEMOGRAPHIC RESEARCH*, vol 16, article 6, pp.141~194.
- Stacy, E.W.(1962), "A Generalized of the Gamma Distributions", *The Annals Mathematics Statistics*, Vol.33, No.3, pp.1187~1192.
- United Nations.(2009), 「World Population Prospects. The 2008 Revision」, New York: United Nations Publication.

<부 록>

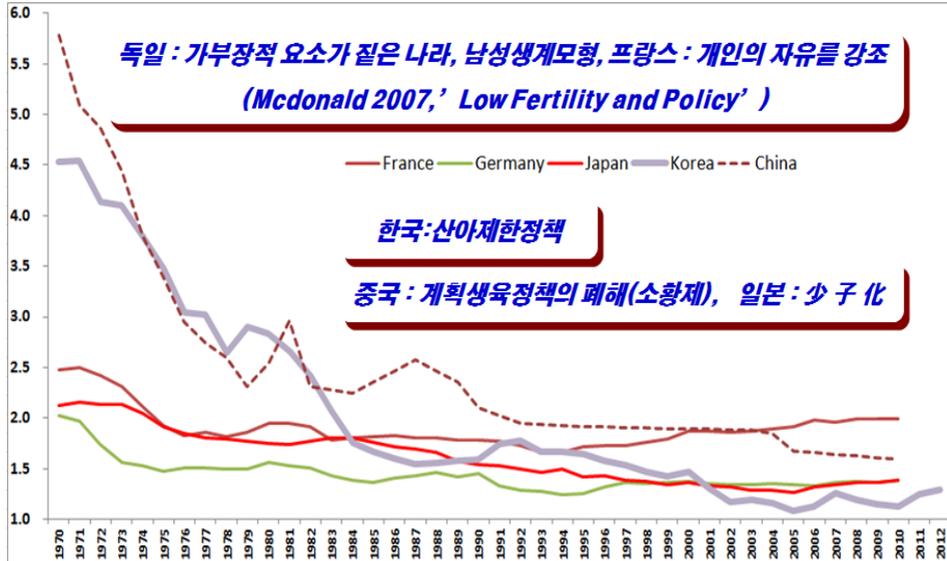
1. 주요 비교 국가별 TFR 추이



[부그림 3-1] 주요 비교 국가별 TFR 추이

독일, 한국, 일본은 인구대체 안전선(safty zone) 아래에 TFR이 위치하고 있어 저출산 국가로 판단

- 합계출산율이 2.1일 때를 인구대체출산율(replacement level fertility)
- 합계출산율이 1.5일 때를 인구대체안전선(safty zone) → 이 안전선은 미래의 노동력을 이진 등의 정책으로 보완할 수 있는 최소 출산율(McDonald 2006, 2007)



[부그림 3-2] 독일, 프랑스, 일본, 중국, 한국 5개국의 출산율 비교

2. 일반로그감마모형(GLG) History

- Stacy(1962) : The generalized gamma distribution, introduced by Stacy Three-parameter distribution(생애(lifetime)분포로 제안)

$$f(x) = \frac{\beta}{\Gamma(k)} \left(\frac{t^{\beta k - 1}}{\alpha^{\beta k}} \right) \exp \left[- \left(\frac{t}{\alpha} \right)^\beta \right], \quad t > 0$$

- Prentice(1974), Lawless(1980) $T = \log T$

$$f(y) = \frac{1}{b\Gamma(k)} \exp \left[k \left(\frac{y-u}{b} \right) - \exp \left\{ \left(\frac{y-u}{b} \right) \right\} \right], \quad -\infty < y < \infty$$

여기서, $u = \log \alpha$, $b = \beta^{-1}$, $(Y-u)/b$ has a log gamma distribution.

문제점 : $k \rightarrow \infty$, $(Y-u)/b$ become infinite.

delta method etc ...

$$W = \sqrt{k} \left(\frac{Y-u}{b} - \log k \right) = \frac{Y-\mu}{\sigma}, \quad \sigma = b/\sqrt{k}, \quad \mu = u + b \log k$$

$$f(w, k) = \frac{k^{k-1/2}}{\Gamma(k)} \exp \left[\sqrt{k} w - k \exp \left\{ \frac{w}{\sqrt{k}} \right\} \right], \quad -\infty < w < \infty$$



□ Farewell and Prentice(1977) $q = 1/\sqrt{k}$

$$f(w, k) = \frac{|q|}{\Gamma(q^{-2})} (q^{-2})^{q^{-2}} \exp[q^{-2}(qw - \exp\{qw\})],$$

$$-\infty < w = \frac{y - \mu}{\sigma} < \infty$$

□ 출산력 모형 GLG

$$f(x) = \frac{C|\lambda|}{b\Gamma(1/\lambda^2)} \left(\frac{1}{\lambda^2}\right)^{\lambda^{-2}} \exp\left[\frac{1}{\lambda}\left(\frac{x - \mu}{b}\right) - \frac{1}{\lambda^2} \exp\left\{\lambda\left(\frac{x - \mu}{b}\right)\right\}\right]$$

3. 시계열 분석 목적과 모형

가. 시계열 분석 목적

예측목적 : 추세분석, 평활법, 분해법, ARIMA

* 평활법 : 시계열을 구성하는 각 성분들을 구분하지 않고 평활에 의해 불규칙성분을 제거 후 미래예측

→ 일정주기 없을 시(지수평활법), 일정주기(Winters의 계절지수 평활법)

* 분해법 : 시계열을 구성하는 각 성분들을 따로 구분한 후 미래예측

이동평균법, 계절조정(Seasonal adjustment)

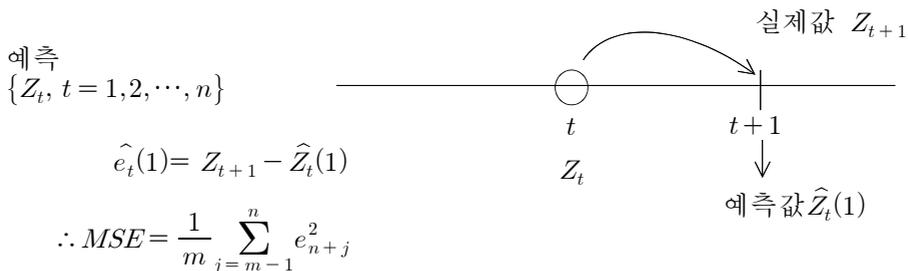
→ 추세에만 관심

시스템 이해 : 스펙트럼, 개입, 전이함수, AR오차모형

(제어목적) → optimal point ?

나. Notation

$Z_t, t = 1, 2, 3, \dots$ 시차(time lag), Z_1, Z_2, Z_3, \dots



※ 회귀모형에 의한 예측

$$\begin{aligned}
 Z_t &= f(X_t; \beta) + \epsilon_t, \quad t = 1, \dots, n & Z_{n+1} &= \beta_0 + \beta_1 X_{n+1,1} + \dots + \beta_p X_{n+1,p} + \epsilon_{n+1} \\
 E(Z_t | X_t) &= f(X_t; \beta) & Z_n(1) &= E(Z_{n+1} | Z_1, \dots, Z_n) \\
 & & \hat{Z}_n(1) &= X'_{n+1} \beta
 \end{aligned}$$

$$\hat{e}_t(1) = Z_{t+1} - \hat{Z}_t(1) \Rightarrow Z_{t+1} = \hat{Z}_t(1) + \hat{e}_t(1)$$

의미 : 새로운 관측값 = 가장 최근의 예측값 + 최근 1시차 후의 예측오차

추정된 예측값의 갱신을 위해 모든 관측값을 저장하고 있을 필요가 없다. 즉, 가장 최근의 추정된 예측값만 저장, 관측값이 추가됨에 따라 새롭게 추정된 예측값을 구할 수 있다.

다. 확률과정 이전과 이후 Notation

확률과정 이전 Notation : $Z_t = \beta_0 + \epsilon_t, \quad Z_t = \beta_0 + \beta_t t^i + \epsilon_t, \dots$

확률과정 이후 Notation :

Z_t : 확률보행과정(random walk process) → 절편이 없는 확률보행과정

$$\begin{cases}
 Z_t = Z_{t-1} + \epsilon_t, & \text{Random walk without Drift} \\
 Z_0 = 0, & t = 1, 2, \dots
 \end{cases}$$

이 과정의 분산 $Var(Z_t) = Var(\sum_{i=1}^t \epsilon_i) = t\sigma_\epsilon^2, r_k = Cov(Z_t, Z_{t+k}) = t\sigma_\epsilon^2,$

확률보행과정은 대표적인 비정상확률과정

Z_t : 확률보행과정(random walk process) → 절편이 있는 확률보행과정

$$Z_t = \delta + Z_{t-1} + \epsilon_t, \quad \text{Random walk with Drift}$$

$$E(Z_t) = t\delta, \quad Var(Z_t) = Var(\sum_{i=1}^t \epsilon_i) = t\sigma_\epsilon^2$$

라. 정상확률과정

이동평균과정	선형과정 = 무한이동평균과정	AR 과정 = $\psi_j = \phi^j$
$\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ $Z_t = \mu + \epsilon_t - \theta\epsilon_{t-1},$ $t = 1, 2, \dots, E(Z_t) = \mu$	$\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ $Z_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$ $E(Z_t) = \mu, \text{Var}(Z_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2,$ $r_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$	$\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ $Z_t - \mu = \phi(Z_{t-1} - \mu) + \epsilon_t$ $Z_t = \delta + \phi Z_{t-1} + \epsilon_t, \delta = (1 - \phi)\mu$ $\dot{Z}_t = \phi \dot{Z}_{t-1} + \epsilon_t, \dot{Z}_t = Z_t - \mu$ $\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$ (1차인 경우)

마. 자기회귀[AR(p)]과정, 이동평균[MA(q)]과정

자기회귀과정(AutoRegression process) : $Z_t = f(Z_{t-1}, Z_{t-2}, \dots) + \epsilon_t$

1차 자기회귀	p차 자기회귀
$Z_t - \mu = \phi(Z_{t-1} - \mu) + \epsilon_t$ $Z_t = \delta + \phi Z_{t-1} + \epsilon_t, \delta = (1 - \phi)\mu$ $\dot{Z}_t = \phi \dot{Z}_{t-1} + \epsilon_t, \dot{Z}_t = Z_t - \mu$ $\dot{Z}_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}$ (1차인 경우)	$Z_t - \mu = \phi_1(Z_{t-1} - \mu) + \dots + \phi_p(Z_{t-p} - \mu) + \epsilon_t$ $\dot{Z}_t = \sum_{j=1}^p \phi_j \dot{Z}_{t-j} + \epsilon_t$ 한편 $Z_t = \delta + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \epsilon_t$ $\delta = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu$

AR(1)

$Z_t = \delta + \phi_1 Z_{t-1} + \epsilon_t$ $\begin{cases} \phi > 0 : \text{변동이 심하지 않음, 부드러움} \\ \phi < 0 : \text{변동이 심함, 반대부호값인점} \end{cases}$ $E(Z_t) = \delta + \phi\mu, \text{Var}(Z_t) = \phi^2 \sigma_z^2 + \sigma_\epsilon^2, \delta = (1 - \phi)\mu$	AR(1)과정 정상조건 : $ \phi < 1$ $\rho_k = r_k / r_o = \phi_k$
---	--



AR(1)과정의 ACF PACF는 어떻게 ?

$$AR(1) = \begin{cases} \phi > 0, ACF : \text{지수적 감소}, PACF : \text{only 1차} + lag \\ \phi < 0, ACF : \text{양, 음 반복} PACF : \text{only 1차} - lag \end{cases}$$

또한 PACF는 양으로 시작하면 $\phi > 0$, 음으로 시작하면 $\phi < 0$

의미 : AR과정의 ACF는 지수적 감소, PACF는 해당차수 이후에는 0 ← 모형식별 활용

$$AR(2) = \begin{cases} \phi_1 > 0, \phi_2 > 0 ACF : \text{지수적 감소}, PACF : \text{only 2차} + lag \\ \phi_1 > 0, \phi_2 < 0 ACF : \text{지수적 감소 후 단절}, PACF : \text{only 2차} + lag, -l \\ \phi_1 < 0, \phi_2 > 0 ACF : \text{양방향 지수적 감소}, PACF : \text{only 2차} - lag, +l \\ \phi_1 < 0, \phi_2 < 0 ACF : \text{양방향 지수적 감소}, PACF : \text{only 2차} - lag \end{cases}$$

참고 : 213p

$$AR(p) = \begin{cases} ACF : \text{지수함수 형태 또는 sin함수 곡선형태로 점차 줄어듦} \\ PACF : \text{시차 } p \text{까지는 0이 아니며 } p \text{차 이후에는 0} \end{cases}$$

$$\text{이동평균과정 } Z_t = \mu + \epsilon_t - \theta\epsilon_{t-1}, \quad \dot{Z}_t = \epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_q\epsilon_{t-q} \rightarrow MA(q)$$

$$t = 1, 2, \dots, E(Z_t) = \mu$$

MA(1)

$Z_t = \mu + \epsilon_t - \theta\epsilon_{t-1},$ $t = 1, 2, \dots, E(Z_t) = \mu$ $\dot{Z}_t = \epsilon_t - \theta_1\epsilon_{t-1}$	$MA(1) = \begin{cases} ACF : \text{Only 1차}, PACF : \text{지수적 감소} \\ \phi > 0 : \text{변동이 심하지 않음, 부드러움} \\ \phi < 0 : \text{변동이 심함, 반대부호 값인점} \end{cases}$ 시차1 이후부터는 0이므로 시차가 2 이상인 시계열들은 서로 상관관계가 없음을 의미하며, 마치 AR(1)의 PACF처럼 시차1 이후에는 0이다.
--	--

AR과정과 MA과정의 쌍대성 (Duality)

MA(1)과정의 PACF=AR(1)과정의 ACF	MA(1)과정의 ACF=AR(1)과정의 PACF
MA(2)과정의 PACF=AR(2)과정의 ACF	MA(2)과정의 ACF=AR(2)과정의 PACF

바. 정상 자기회귀이동평균[ARMA(p,q)]과정

AR과정과 MA과정의 쌍대성 (Duality)

- ① 유한차수 정상 AR과정은 무한차수 MA과정으로 표현 가능, 반대도 가능
- ② 유한차수의 AR과정의 ACF와 유한차수의 MA과정의 PACF는 지수적으로 감소하는 형태

유한차수의 AR과정의 PACF와 유한차수의 MA과정의 ACF는 절단형태를 나타냄

$ARMA(p, q)$:

$$\begin{aligned} Z_t &= \delta + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \\ Z_t - \mu &= \phi_1 (Z_{t-1} - \mu) + \phi_2 (Z_{t-2} - \mu) + \dots \\ &\quad + \phi_p (Z_{t-p} - \mu) + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \end{aligned}$$

사. 비정상(nonstationary) 자기회귀이동평균과정

- 1) 시간의 흐름에 따라 시계열의 관측값이 추세를 가지고 증가
- 2) 산포의 정도가 변하는 경우
- 3) 추세를 가지고 증가하지 않더라도, 국지적으로 그 수준이 같거나 산포의 정도가 동일하나 다른 시간대에서는 서로 수준이 다르거나 산포의 정도가 달라지는 경우
- 4) 계절적인 요인에 의해 시계열이 주기적으로 반복되는 경우

비정상성의 중요한 요인의 추세

- 1) 결정적 추세(deterministic trend) : 추세가 결정적이고 동시에 영원히 지속
- 2) 확률적 추세(stochastic trend) : 인접자료들 간에 강한 양의 상관관계 때문에 어떤 추세가 있는 것처럼 보이는 경우

시계열자료의 분석과 관련된 대부분의 이론들이 정상성을 가정하고 있으므로 시계열이 비정상성을 보이는 경우에는 로그변환 또는 차분등의 적절한 방법에 의하여 시계열이 정상이 되도록 한 후 정상시계열분석의 이론을 적용

아. 정상화 방법

- 1) 분산이 일정하지 않은 경우 : 분산안정화변환(Variance stabilizing transformation)
→ 로그변환이나, 제곱근변환, 로그변환($\ln Z_t$)
- 2) 평균이 일정하지 않은 경우 $Y_t(\log Z_t) = T_t + S_t + I_t$
→ 비정상시계열 $\{Z_t\}$ 이 분산안정화변환 $Y_t = T(Z_t)$ 을 통해서 각 시점에서 동일한 분산을 가진다고 하자. 그러나 변환된 Y_t 의 평균수준이 일정하지 않고 추세를 가진다면 추세를 제거하여 정상화해야 한다.



① 결정적추세와 분해법

$$Y_t(\log Z_t) = T_t + S_t + I_t$$

$$\begin{cases} T_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^k \\ S_t = \sum_{t=1}^m (\beta_{1i} \sin f_i t + \beta_{2i} \cos f_i t) \end{cases} \quad \begin{cases} E(Y_t) = E(T_t + S_t + I_t) = T_t + S_t \\ Var(Y_t) = Var(I_t) = \sigma_I^2 \end{cases}$$

② 확률적추세와 차분법

다른 시간대에서는 시계열의 움직임이 다르지만 본질적으로 시계열의 움직임이 같은 경우가 있으며 이러한 형태의 비정상성을 동질적인 비정상성(homogeneous nonstationarity)이라 한다.

→ 시계열이 확률적추세를 갖는 비정상시계열은 차분을 함으로서 정상시계열로 바꿀 수 있다.

차분연산자(differencing operator) $\nabla : \nabla Y_t = Y_t - Y_{t-1}, \nabla^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$

일반적으로 차수가 d인 다항식추세(polunomial trend)를 가지는 비정상시계열

$$Y_t = T_t + I_t = (\beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_d t^d) + I_t$$

→ 정상시계열을 얻기 위해서는 d차 차분을 해야함

자. 해결방안

※ 추세를 가지는 비정상시계열의 경우

→ 이 추세가 결정적 추세인지 확률적 추세인지를 결정한 후

Unit root test(Dickey와 Fuller) $\begin{cases} \text{확률적 추세 : 차분} \\ \text{결정적 추세 : 추세항포함 모형적합} \end{cases}$

※ 계절성분의 경우에도 결정적인지 확률적인지를 판단하기 위해서는 계절형 단위근검정(seasonal unit root test) 등을 이용한다.

확률보행과정(비정상시계열)

$$Z_t = Z_{t-1} + \epsilon_t, \quad E[Z_t | Z_{t-1}, Z_{t-2}, \dots] = Z_{t-1} \text{---- (1)}$$

$$Z_t = 2Z_{t-1} - Z_{t-2} + \epsilon_t, \quad E[Z_t | Z_{t-1}, Z_{t-2}, \dots] = 2Z_{t-1} - Z_{t-2} \text{---- (2)}$$

(1)은 1차 차분, (2)는 2차 차분으로 정상시계열로 분석



자기회귀누적이동평균과정

d차 차분된 $W_t = (1-B)^d Z_t$ 가 평균수준이 μ 인 정상 $ARMA(p,q)$ 과정을 따를 때 $\{Z_t\}$ 는 자기회귀누적이동평균과정, $ARIMA(p,d,q)$ 과정을 따른다고 한다.

□ 차분이 필요한 경우

1) 시계열 그림을 관측하여 추세를 가지고 있는가를 확인

추세가 결정적이라면 추세항포함시켜 모형적합

추세가 확률적이라면 차분에 의해 정상화

→ 결정적 추세가 아니면서 몇 개의 추세가 연속해서 나타나는 비정상시계열이 있다면 차분 (1-B)를 취하여 ARIMA모형으로 분석을 시도하는 것이 좋은 방법이다.

$Z_t = Z_{t-1} + \epsilon_t$, $E[Z_t | Z_{t-1}, Z_{t-2}, \dots] = Z_{t-1}$ 의 경우 ARIMA(0,1,0)로 표기할 수 있다.

ARIMA(p,d,q)모형에서 p=0인 경우를 IMA(d,q)모형이라 부른다.

$$\rightarrow Z_t = Z_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$$

이는 정상 ARMA(1,1)모형의 형태를 갖지만 비정상성을 내포하는 IMA(1,1)=ARIMA(0,1,1)

$$Z_t = Z_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$$

ARIMA(p,d,q)모형에서 q=0인 경우를 ARI(p,d)모형이라 부른다. ARI(1,1)=ARIMA(1,1,0)

$$Z_t = (\phi + 1)Z_{t-1} - \phi Z_{t-2} + \epsilon_t$$

→ 혹, AR(2) 모형 같지만 unit root test 실시 후 ARI(1,1)모형이라 부름

2) 천천히 감소하는 SACF를 가지는 시계열자료는 비정상성을 가지고 있을 가능성이 높으며 따라서 ARIMA모형을 적용한다.

□ 과대차분(Overdifferencing)

차분 후에 시계열의 분산이 차분 전의 시계열의 분산보다 더 커지면 과대차분일 가능성이 높다고 보아 차분을 하지 않는 것이 좋다.

□ Unit Test

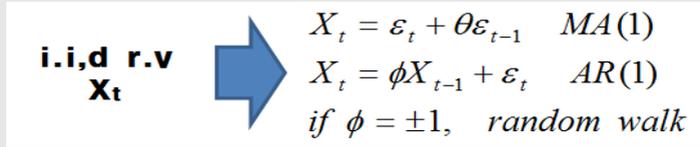
$$X_t = X_{t-1} + \varepsilon_t \quad r.w$$

$$\left(X_t = \phi X_{t-1} + \varepsilon_t \right)$$

$$\left(\phi = \pm 1 \right)$$

$$\left[\begin{array}{l} X_t = \varepsilon_t + \theta \varepsilon_{t-1} \quad MA(1) \\ X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t \quad AR(p) \end{array} \right]$$

ϕ 가 ± 1 인지 아닌지를 검정하는 test : Unit test
 root test, Dickey Feller, ADF(Augmented Dickey Feller) 검정



$$X_t - \mu - \phi_1(X_{t-1} - \mu) - \phi_2(X_{t-2} - \mu) - \dots - \phi_p(X_{t-p} - \mu) = \varepsilon_t$$

$$\Rightarrow X_t = \mu + \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t$$

$$\underline{X_t = \phi_1 X_{t-1} + \varepsilon_t}$$