

# 한국의 강력 범죄 발생 추이 및 통제 요인 연구<sup>†</sup>

권태연<sup>1</sup> · 전새봄<sup>2</sup>

<sup>1</sup>고려대학교 경제연구소 · <sup>2</sup>고려대학교 통계학과

접수 2016년 9월 21일, 수정 2016년 10월 31일, 게재확정 2016년 11월 5일

## 요약

우리나라의 5대 강력범죄 (살인, 강도, 강간, 폭력, 절도) 발생의 증가추세는 우리나라의 사회, 경제적 요인의 변화 추세와 무관하지 않으며, 이와 관련한 논의는 여러 사회과학 연구에서 논의되어져 왔으나 시계열 자료의 특성을 제대로 반영하지 않은 경우가 많다. 이에 본 연구에서는 강력범죄 변화의 추이를 살펴보고 그 통제 요인들에 관하여 논의하였다. 통제 요인들을 살펴봄에 있어 시간, 계절 및 순환과 같이 시계열 자료로써 갖는 내재적 요인들과 경제적, 사회변동 및 범죄통제에 관련한 외재적 요인들로 범주화 하여 고려했다. 또한 시계열자료가 본질적으로 갖는 자기상관성을 반영한 모형 역시 고려하여 비교하였다. 이러한 다양한 시계열 모형들을 통하여 5대 강력범죄의 발생요인을 점검하는 한편 발생건수를 예측함으로써 강력 범죄에 대한 예방적 정책적 도움을 주고자 하였다.

주요용어: 강력범죄, 내적요인, 시계열 회귀모형, 외적요인, 자기상관성.

## 1. 머리말

우리나라 5대 강력범죄는 살인, 강도, 강간, 폭력 그리고 절도를 의미하며, 이러한 강력 범죄는 점점 증가 추세에 있다 (Ju와 Cho, 2012; Park과 Kim, 2015). 강력 범죄 발생 건수의 증가는 많은 피해자를 양산할 뿐 아니라, 인터넷과 미디어의 발달로 인해 국민들의 심리적 동요와 불안정을 심화시켜 사회적 안정을 저해하므로 강력 범죄 증가를 억제하기 위한 노력 및 대책 마련이 절실하다. 그러나 범죄에 대한 대책 및 예방 정책 수립을 위해서는 범죄 발생에 대한 현황 및 추이, 이러한 강력 범죄와 관련된 요인 등에 대한 파악이 우선되어야 할 것이다.

범죄의 현황 및 실태, 그리고 그에 대한 원인은 한국에서도 강력범죄 뿐 아니라 청소년 범죄에 이르기까지 상당히 연구되어 왔다. Park (2002)은 실업률 등의 경제 지표가 우리나라 범죄율 증가에 미치는 영향을 확인하였으며, Kim (2004)은 도시 내 녹지 환경의 부족이 범죄에 미치는 영향을 연구하였다. Lee (2010)는 시간·요일별·연령별·성별·교육수준별 범죄 현황을 통해 이들의 연관성을 연구하였으며, 이 외에도 국내 총소득, 실업률, 지니계수 등과 같은 경제적 요인, 경찰 1인당 인구변수 등과 같은 범죄통제 요인 등 다양한 범죄 발생 원인이 제시되어 왔다 (Park 등, 2009).

그러나 범죄 발생을 단순히 현황 측면에서 살피는 것이 아니라, 시간 흐름에 따른 발생 추이 및 변동 요인에 대해 심층적으로 살피기 위해서는 기존의 횡단면적 연구는 방법론적으로 한계가 있으며, 따라서 이러한 결과에 대한 해석 역시 바로 적용하여 예방 정책을 논의하는 것은 적절하지 않다. 범죄 발생 증

<sup>†</sup> 이 논문은 고려대학교 재원을 일부 지원받아 수행된 연구임.

<sup>1</sup> (02841) 서울시 성북구 안암로 145, 고려대학교 경제연구소, 연구교수.

<sup>2</sup> 교신저자: (02841) 서울시 성북구 안암로 145, 고려대학교 통계학과, 연구교수.

E-mail: alwaysns@korea.ac.kr

가 추세 및 변동은 내재적으로 시간에 대해 자기상관성을 갖고 있으므로 일반적인 회귀모형을 근간으로 하는 방법이 아닌 시계열적인 접근이 필요하다.

범죄 발생 연구와 관련한 시계열 연구는 외국의 경우 상당히 연구되어 왔으며, Corman 등 (1987)과 Corman 과 Mocan (2000)은 뉴욕의 범죄발생에 대하여, Shah Habibullah and Baharom (2009)은 말레이시아의 범죄 발생, Deadman과 Pyle (1997)는 영국의 범죄 발생에 대해, 그리고 Beki 등 (1999)는 약 50년에 걸친 장기간의 네덜란드 내 범죄율 변동 및 범죄 발생 요인과의 관계를 파악하였다. 이처럼 외국에서는 다양한 시계열 모형을 이용하여 주요 국가 및 대도시의 범죄 발생의 사회, 경제, 범죄 통제적 요인을 파악하고 있는데 비해, 우리나라의 경우는 이러한 시계열 접근은 미비한 현실이다. Park 등 (2009)이 자기상관 오차항을 갖는 회귀모형을 사용하여 한국의 범죄율 관련 요인 연구를 시도한 바 있으나, 최근 다양한 공공 자료의 공개로 인해 자료 접근의 한계가 극복되었으므로 이러한 자료들의 특성에 맞는 다양한 범죄 발생 연구는 매우 필요하다.

이에 본 논문에서는 최근 한국에서 발생하는 주요 5대 강력 범죄 추이를 살펴보고, 월별 범죄 발생 건수의 특성을 반영할 수 있는 ARIMA모형, 시계열 회귀모형, 자기 회귀오차항 모형을 통해 범죄발생 요인에 대한 영향력을 파악하는 한편, 향후 범죄 발생에 대한 예측을 실시하여 모형들을 비교하였다. 이때 범죄 발생 요인을 크게 시계열 자료로써의 내재적 요인 (endogenous factor)과 외재적 요인 (exogenous factor)으로 나누어 살펴보았다.

본 논문에서 정의한 범죄 발생의 시계열적 내재적 요인은 범죄 발생과 관련된 시계열 자료 자체에서 파악될 수 있는 요인으로 추세 (trend), 계절 (seasonality) 및 순환(cycle) 요인을 의미한다. 본 논문에서는 범죄발생에 대한 추세요인을 반영하기 위해 선형추세의 시간변수 (time variable)와 시차변수 (lagged variable)를 사용하였으며, 계절성을 반영하기 위해 월별 더미변수 (month dummy variable)를 이용하였다. 순환요인을 반영하기 위해서는 삼각함수 (trigometric function)를 이용하였다.

범죄 발생의 외재적 요인은 범죄 발생에 영향을 주는 외부의 독립변수들 (independent variable)로 거시적 관점으로 크게 경제변동, 사회변동 그리고 범죄통제요인으로 나누고 이들과의 관계를 다양한 시계열모형으로 살펴보았다. 우선 경제변동 요인으로는 GDP와 실업률을 고려하였고, 사회변동 요인으로는 외국인 유입수와 젊은 남성의 비율변화를 살펴보았으며, 범죄통제 변인으로 경찰 1인당 담당인구수와 검거자수를 고려하였다.

본 논문은 범죄 발생 증가 추세에 영향을 미치는 다양한 요인들의 시계열 자료를 이용하여 자료의 성격에 부합하는 모형들의 적합도 및 예측력을 비교하였다는 점에서 기존 연구와 차별되며, 향후 한국의 범죄 관련 연구 방법론적 확장에 있어서도 의의가 있다고 할 수 있겠다.

본 논문은 총 4절으로 구성되어 있으며, 2절은 범죄발생의 요인에 대한 기존 연구 결과들을 근거로 본 논문의 연구 자료인 5대 강력범죄 발생 건수 및 고려할 요인 변수들에 대해 기술하고 있다. 3절은 본 논문에서 고려하는 시계열 모형을 소개하고, 4절에서는 범죄 발생 건수 시계열 회귀모형들의 분석 결과를 제시, 비교하였다. 마지막으로 5절은 결론 및 제언을 제시하였다.

## 2. 강력 범죄 발생 및 범죄 발생의 요인

본 연구에서 지칭하는 강력 범죄 발생 건수는 주요한 5대 강력 범죄인 살인, 강도, 강간, 폭력, 절도 발생건수를 의미하며, 분석에 사용되는 자료는 대검찰청 범죄통계시스템에서 제공하는 2003년 1월부터 2013년 12월까지 10년간의 월별 강력범죄 발생건수 ( $y_t$ )이다. 최근 들어 추가로 공개된 2013년 이후 자료는 범죄에 대한 정의 및 범주 구분이 이전 자료와 차이가 있어 범죄발생에 대한 시계열 자료의 내적 일치성을 위하여 분석에 제외하고 2013년까지의 자료만을 이용하였다. 범죄 발생의 현상 및 범죄 발생 요인을 보다 잘 파악함으로써 미래의 범죄발생에 대한 예방적 정책적 도움을 주기 위해서, 본 연구는 범

죄 발생 건수 자체에 대한 시계열적 접근으로 설명하는 모형과 외부적인 위험요인을 고려하는 회귀적 접근을 모두 고려하고, 이를 각각 내재적 요인과 외재적 요인으로 명명하여 비교 분석하고 있다.

한 사회의 범죄 발생이 그 사회의 경제적, 사회적, 정치적 여러 특징에 의해 영향을 받으므로 기존의 횡단면적 연구인 회귀모형은 범죄 발생과 관련한 기저의 관련 원인을 찾는 심층적 연구가 가능하다는 장점이 있는 반면, 미래의 범죄발생 예측을 위해서는 관련 외부요인 즉 독립변인에 대한 예측이 선행되어야 한다는 점에서 예측모형으로 활용되기에 어려움이 존재한다. 반면, 내재적 요인만을 고려하는 경우 이러한 외부요인에 대한 예측이 필요없이 기존의 강력 범죄 발생의 추이 및 패턴을 통해 향후의 범죄 발생을 예측한다는 점에서 그 장점을 찾을 수 있다. 그러나, 내재적 요인을 이용한 시계열 모형은 향후의 범죄 발생이 기존의 범죄 발생 추이나 패턴과 다른 양상을 갖는 경우 예측력이 떨어지게 되므로, 범죄 발생을 예측함에 있어 내재적 접근 및 외재적 접근 둘 다 살펴볼 필요가 있다고 하겠다.

### 2.1. 범죄 발생의 내재적 요인

범죄 발생을 설명하기 위한 시계열적 내재적 요인은 범죄발생을 측정된 자료가 시계열 자료로써 갖는 특징을 설명하기 위한 요인으로 시간, 계절성 및 순환요인을 의미한다. Figure 2.1에 제시된 바와 같이, 2003년 1월부터 2013년 12월까지 우리나라 5대 강력 범죄의 발생 건수의 추이를 살펴본 결과, 5대 강력범죄의 발생건수는 전반적으로 선형적으로 증가하는 추세를 따르고 있으며, 분산은 큰 변화 없음을 알 수 있다. 또한 매 해 1, 2월엔 범죄 발생 건수가 저점을 보이는 반복적인 양상을 갖고 있어, 강력 범죄의 발생이 계절적 특성을 가짐을 알 수 있다. 이에 시간, 계절성 및 순환요인을 고려한 시계열 회귀모형을 통해 범죄 발생의 내재적 특성을 반영을 고려할 수 있다 (Park과 Kim, 2002).

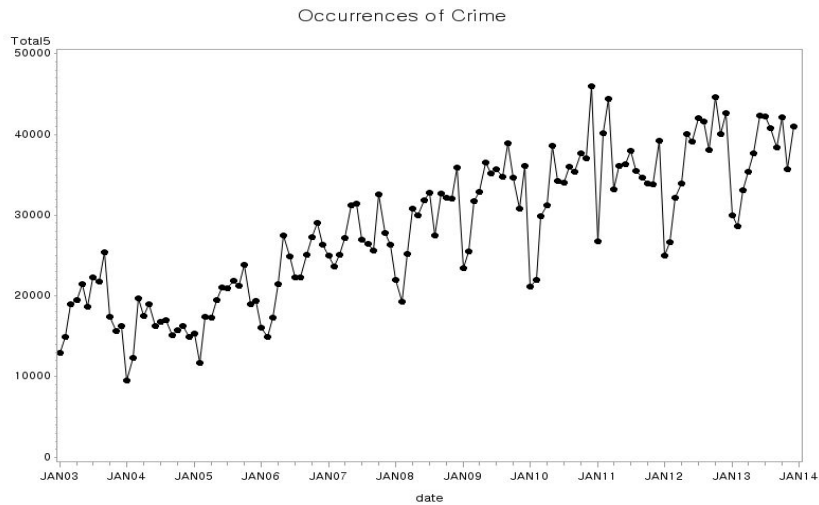


Figure 2.1 Time plot of 5-violent crime outbreak

시간추세를 반영하기 위해 선형추세 (linear trend)인 추세변수  $TR_t = t$  를 고려하고, 계절성을 반영하기 위한 가변수로 범죄 발생건수의 시차변수 ( $y_{t-1:t-13}$ )를, 월별 효과를 반영하기 위한 월별 더미변수 ( $m_{1:11}$ )를 모형에 고려하였으며, 순환요인은 고정계절변동을 위한 삼각함수 (trigometric function)인  $\sin(2\pi t/12)$ ,  $\cos(2\pi t/12)$  과 확산계절변동을 위한 삼각함수  $t \sin(2\pi t/12)$ ,  $t \cos(2\pi t/12)$  를 모형에 고려하였다.

## 2.2. 범죄 발생의 외재적 요인

한 사회 내의 범죄발생은 경제적 요인, 인구 사회적 변동 요인, 범죄통제 요인 등 다양한 요인에 의해 영향을 받는다. 먼저 범죄 발생을 산업화, 경제발전, 경제적불평등 및 경제적 빈곤과 같은 경제적 요인과 관련하여 파악하는 연구들은 다음과 같다. Corman and Mocan (2000)은 미국 뉴욕시 범죄발생을 국가로부터 생계보조금을 받는 가정의 수 (number of case to aid to families with dependent children)를 고려함으로써 경제적빈곤이 특정 범죄의 발생과 관련이 있음을 설명하였으며, Corman 등 (1987)은 뉴욕의 실업률과 범죄의 발생의 관계를 설명하였다. Deadman and Pyle (1997)은 국내총생산 (GDP), 실업률, 국내총소비를 범죄의 경제적 요인으로 고려하였다. Shah Habibullah and Baharom (2009)는 말레이시아 범죄발생의 경제적 원인으로 국민총생산 (GNP)를, 그리고 Beki 등 (1999)는 가계지출 (personal consumption of house holds per capita)을 이용하여 분석하였다. 국내 연구로는 Park 등 (2009) 및 Park (2002)에서 국내총소득, 실업률 및 지니계수등의 경제적 범죄요인을 밝힌 바 있다. 이에 본 논문에서는 범죄의 경제적 요인으로 국내총생산 (GDP)을 고려함으로써, 전반적인 사회의 경제규모 및 경기를 파악하여 전반적인 경제상황이 범죄의 원인이 되는지를 파악하였다. 또한, 실직 등을 이유로 경제활동을 하지 않는 집단은 빈곤이나 상대적 박탈감을 느끼는 환경에 보다 더 노출되어 있기 때문에 범죄행위를 저지를 가능성이 크다고 알려져 있으므로 비경제활동 인구수의 증가를 5대 강력범죄 발생 증가에 유의한 영향을 미치는 변수로 고려하였다.

두 번째 범죄발생요인으로 인구 변동, 도시화 등 환경 변동을 포함한 사회변동 요인을 들 수 있다. Kim (2004)은 도시 내 녹지 환경의 부족과 범죄의 관계를 연구한 바 있으며, Corman 등 (1987)은 16세에서 24세의 젊은 남성 인구의 비율 변화와 범죄에 미치는 영향을 연구한 바 있다. Beki 등 (1999) 역시 남성 인구의 연령별 구조 변화와 범죄율의 연관관계에 대하여 파악하였다. 이들은 강력범죄자의 주요 인구통계학적 특징이 젊은 남성임을 보이고 있으므로, 전체인구에서의 젊은 남성의 비율변화는 강력범죄 발생과 관련한 외재적 사회 변동 요인으로 고려할 수 있다. 한편 체류 외국인의 급격한 증가는 여러 가지 사회적 문제점을 발생시킬 가능성이 크다는 것이 다양한 연구를 통해 검증되었다 (Choi, 2007; Lee 2007). 2003년 대비 2007년에는 체류 외국인이 1.7배 증가하는 동안 외국인 범죄는 2.36배 증가하였으며, 외국인 범죄자 중 16%는 범죄를 목적으로 입국했다는 보도와 함께 실제로 국내에 입국한 외국인이 범죄를 일으키는 건수는 점점 증가하는 추세이다. 이러한 맥락에서 외국인 수 역시 범죄 발생 건수에 영향을 끼치는 사회변동 요인으로 고려할 수 있다. 이에 본 논문에서는 범죄 발생에 대한 사회변동 요인으로 외국인 유입수와 전체 인구 대비 젊은 남성의 비율 변화를 고려하였다.

세 번째 범죄발생요인은 앞의 두 요인에 비해 범죄 발생에 보다 직접적 요인인 범죄통제 요인으로, 경찰규모의 변화 (Deadman and Pyle, 1997; Corman 등, 1987; Park 등, 2009)와 검거율 (Beki 등, 1999; Corman and Mocan, 2000) 등을 들 수 있다. 경찰의 규모가 클수록 검거율이 높을수록 범죄 행위를 하는데 있어 심리적 압박감을 받을 가능성이 크다. 따라서 낮은 치안율이 범죄 발생 건수의 증가에 영향을 미치는 것으로 볼 수 있다. 이런 점을 고려하여 본 논문에서는 경찰 1인당 담당인구수와 5대 강력범죄 검거자수, 전일 검거율을 범죄 발생 건수에 영향을 미치는 변수로 고려하였다.

이러한 강력범죄 발생의 외재적 요인인 경제적, 사회변동 및 범죄통제 요인에 대한 2003년 1월부터 2013년 12월까지의 양상은 Figure 2.2과 같다. 이때 경제적 요인인 국내총생산 (GDP)은 통계청에서 제공하는 동 기간 분기별 자료를 월별자료로 변환하여 이용하였으며, 실업자 수 (unemp)의 경우는 통계청에서 제공하는 동 기간 월별 자료를 이용하였다. 사회변동 요인인 외국인 유입수 (foreigner)와 16~24세의 젊은 남성 (r\_male)의 비율변화는 통계청에서 제공하는 동 기간의 인구 자료를 월별 추출하여 사용하였고, 마지막으로 범죄통제 요인인 검거자수 (arrest)와 경찰 1인당 담당인구수 (police)는 대검찰청에서 제공하는 2003년부터 2013년까지 연도별 자료를 선형 보간법을 이용하여 월별자료로 변환하여 분석에 사용하였다.

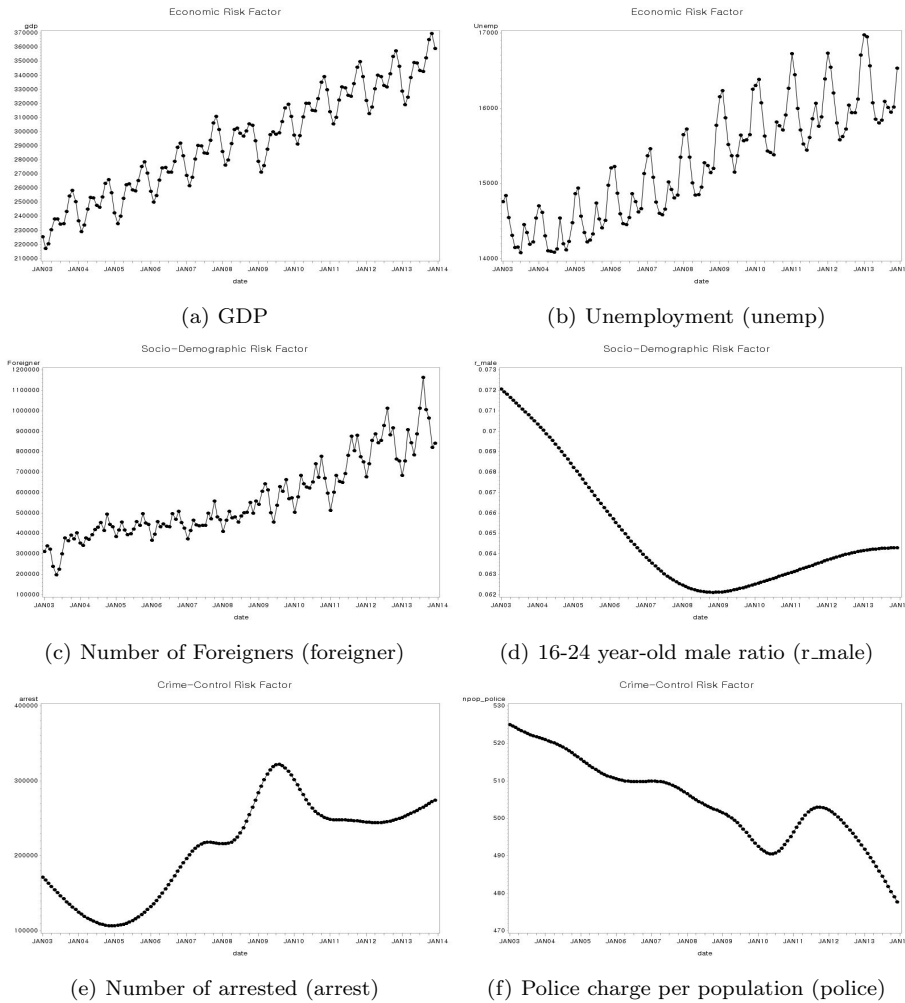


Figure 2.2 Time plot of risk factors of crime

국내총생산의 분기별 효과는 뚜렷하게 나타나며 분석기간 동안의 분산의 변화는 미미함을 확인하였다. 비경제활동 인구수는 다른 자료에 비해 강한 계절성을 보이면서 분산이 점차 증가하고 있으며, 상승하는 추세를 가지고 있다. 외국인 유입자수의 경우, 분산이 증가하는 폭이 점차적으로 커지며 꾸준히 상승하는 추세와 계절성이 있다.

한편, 16~24세 남성 인구는 2008년에서 2009년 사이 추세의 방향이 전환되는 시점이 있음을 확인할 수 있었다. 이에 추세의 구조변화를 검정하는 Chow Test (Chow, 1960)을 통하여 구조변화 여부와 그 시점을 테스트한 결과, 2008년 6월 ( $F$ -value: 2.88,  $p$ -value: 0.0384) 추세의 구조변화가 있음이 파악되어 이러한 구조변화 시점에 대한 더미 변수 (Date.Change)와 남성 인구 비율과 구조변화 시점의 교호 효과를 추가로 모형 내에 고려하였다. 검거자수는 2005년까지 감소하는 추세에서 전환되어 2010년까지 증가추세에 있었으나, 2011년 이후 큰 추세변화 없이 일정 수준으로 유지됨을 볼 수 있다. 경찰 1인당 담당인구수는 전반적으로 감소추세에 있으나 2012년에 소폭 증가하였다.

### 3. 범죄 발생 요인 파악을 위한 시계열 모형

본 연구에서 적용한 시계열 모형은 ARIMA 모형, 회귀모형, 자기회귀 오차항 회귀모형의 세 가지 범주로 나누어 볼 수 있으며, 시계열 회귀모형과 자기회귀 오차항 회귀모형을 위해 범죄발생건수의 시계열적 시간 효과를 반영할 수 있는 추세변수와 시차변수를 반영한 모형, 여기에 계절효과와 반영을 위한 월별 효과를 고려한 모형, 순환효과까지 추가적으로 고려한 모형을 적용하였다. 그 다음 이러한 범죄발생의 내재적 변동요인 이외에, 범죄발생에 영향을 끼치는 요인으로 알려진 경제변동, 사회변동 그리고 범죄통제요인들 및 이들의 구조적 변화효과를 함께 고려한 모형을 적용하였다 (Ryu와 Kim, 2013). 본 연구에서 고려한 상기의 총 13개 시계열 모형은 다음의 Table 3.1에 정리하였다.

우선 내재적 요인에 대한 모형 (Endogenous Model)으로는  $ARIMA(p, d, q)$ 에 덧붙여 3개 유형의 시계열 모형으로 (1) Linear Trend, 즉 시간효과  $t$ 와 시차변수  $y_{t-1}, \dots, y_{t-12}$  만을 고려하는 추세 모형, (2) Linear Trend and Month effect, 즉 추세뿐 아니라 계절성을 반영하기 위해 월별 가변수  $D_1, \dots, D_{11}$ 을 고려하는 모형, (3) Linear Trend, Month effect and Cycle effect, 즉 추세와 계절성뿐 아니라 확산주기까지 반영하기 위해 고정계절 변동을 나타내는 삼각함수  $\sin(2\pi t/12)$ ,  $\cos(2\pi t/12)$  및 확산계절 변동을 나타내는 삼각함수  $t \sin(2\pi t/12)$ ,  $t \cos(2\pi t/12)$ 항들을 포함하는 모형을 고려하였다.

**Table 3.1** Time series models for crime outbreaks

| Factors          | Model  | Model Formula   |
|------------------|--|---|
|                  | $ARIMA(p, d, q)$   | $(1 - \phi_p(B))(1 - \Phi_{ps}(B))\nabla_D \nabla_d y_t = (1 - \theta_q(B))(1 - \Theta_{qs}(B))\epsilon_t,$<br>$\epsilon_t \sim N(0, \sigma^2)$   |
| Endogenous Model | (1) Trend  | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$  |
|                  | (2) Trend and Month effect                               | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \sum_{m=1}^{11} \beta_{3m} D_m + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$   |
|                  | (3) Trend, Month and Cycle effect                        | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \beta_{41} \sin(\frac{2\pi t}{12}) + \beta_{42} \cos(\frac{2\pi t}{12})$<br>$+ \beta_{43} t \sin(\frac{2\pi t}{12}) + \beta_{44} t \cos(\frac{2\pi t}{12}) + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$   |
| Exogenous Model  | (4) Trend with Exogenous factors                         | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \sum_{p=1}^6 \beta_{5p} x_{tp} + \beta_6 C_{t*} + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$  |
|                  | (5) Trend and Month effect with Exogenous factors        | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \sum_{m=1}^{11} \beta_{3m} D_m + \sum_{p=1}^6 \beta_{5p} x_{tp} + \beta_6 C_{t*} + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$   |
|                  | (6) Trend, Month and Cycle effect with Exogenous factors | $y_t = \beta_0 + \beta_1 t + \sum_{l=1}^{12} \beta_{2l} y_{t-l} + \beta_{41} \sin(\frac{2\pi t}{12}) + \beta_{42} \cos(\frac{2\pi t}{12})$<br>$+ \beta_{43} t \sin(\frac{2\pi t}{12}) + \beta_{44} t \cos(\frac{2\pi t}{12}) + \sum_{p=1}^6 \beta_{5p} x_{tp} + \beta_6 C_{t*} + \epsilon_t,$<br>$\epsilon \sim N(0, \sigma^2),$<br>$\epsilon_t = \sum_{p=1}^{p^*} \phi_p \epsilon_{t-p} + a_t, a_t \sim N(0, \sigma^2),$ |

다음으로 외재적 요인에 대한 모형 (Exogenous Model)으로는, 상기의 여러 내재적 요인 뿐 아니라 범죄발생과 관련 있는 다른 외재적 요인인 외부위험요인  $x_{t1}, \dots, x_{t6}$  및 이와 관련한 구조변화  $C_{t*}$ 를 추가적으로 반영하는 3개 유형의 시계열 모형 (4)~(6)을 고려하였다. 이때 외부위험요인  $x_{t1}, \dots, x_{t6}$ 는 경제적 요인, 인구 사회적 변동 요인, 범죄통제 요인을 대표하는 국내총생산 (GDP), 실업자 수 (unemp), 외국인 유입수 (foreigner), 16~24세의 젊은 남성 (r.male)의 비율, 검거자수 (ar-

rest)와 경찰 1인당 담당인구수 (police)를 의미하며,  $C_{t*}$ 는 16~24세 남성 인구비 시계열에서 2008년 6월부터 나타나는 구조변화를 반영하기 위한 변수로,  $C_{t*}$ 는 2008년 6월부터는 1의 값을 갖고 그 이전에는 0의 값을 갖는다.

이러한 내재적 요인 및 외재적 요인에 대한 6개 유형의 모형 (1)~(6)은, 오차항에 대한 가정에 따라 시계열 회귀모형 식 (3.1a)부터 (3.6a), 그리고 자기회귀 오차항을 갖는 시계열 회귀모형 (Lee, 2015; Lim, 2014) 식 (3.1b)부터 (3.6b)로 구분할 수 있다.

#### 4. 범죄 발생에 대한 시계열 모형들의 적합 및 예측결과 비교

##### 4.1. 강력범죄 발생 건수에 대한 ARIMA 모형

첫 번째 모형인  $ARIMA(p, d, q)$  모형 적합을 위해서는 자료의 추세와 분산의 정상화 과정이 필요하다. 예측의 비교를 위하여 모형구축 및 적합은 2012년 12월까지의 자료만을 이용하였으며, 해당 기간 범죄 발생 건수는 Dickey-Fuller 추세 검정 결과 추세 안정화를 위해 1차 차분 및 계절 차분이 필요함을 알 수 있었다. 이에 시계열 분석을 위한 정상화 자료인  $\nabla_{12}\nabla_1$  5대 강력범죄발생건수 ( $\nabla_{12}\nabla_1 y_t$ )를 생성하여 적합하였다. 그 결과  $ARIMA(2, 1, 0)(0, 1, 1)_{12}$ 로 식별되었으며 최종 모형의 모수추정 결과는 다음의 식 (4.1)과 같았다. 모형의 적합도는 SBC값이 2065.56, AIC값은 2057.54였으며 예측치의 RMSE는 4228.24였다.

$$(1 + 0.46B + 0.29B^2)\nabla_{12}\nabla_1 y_t = (1 - 0.59B^{12})a_t. \quad (4.1)$$

##### 4.2. 강력범죄 발생 건수에 대한 내재적 및 외재적 요인 시계열 모형

Table 3.1에 제시된 총 12개의 시계열 모형, 즉 시계열 회귀모형과 오차항이 자기상관을 갖는 회귀모형 식 (3.1a)부터 (3.6b) 적합 시 변수 선택은 모형의 간명성을 위하여 후진소거법 (Backward elimination)을 이용하였으며, 유의한 자기상관성을 갖는 오차항은 Godfrey's Serial Correlation Test (Godfrey, 1978)를 이용하여 AR(13)까지 고려한 후, 후진소거법으로 선택하였다 (Park과 Song, 1998). 한편 각 모형에 대한 예측력 비교를 위하여 분석 자료의 가장 최근 1년치 자료를 제외한 2012년 12월까지의 자료만을 이용하여 모형을 적합하고 그 이후의 자료값을 이용하여 모형의 예측력을 비교하였다.

Table 4.1은 내재적 요인만을 이용한 시계열 모형들 (Table 3.1의 (1)~(3)번 유형), 즉 시계열 회귀모형과 오차항의 자기상관을 갖는 회귀모형의 모형적합 결과이다. 각 모형 적합 시 통계적으로 유의하지 않은 변수들은 후진소거법으로 제거되었으므로 Table 4.1에 제시된 회귀계수 값은 모두 통계적으로 의미 있는 값이므로 표에 대한 간명성과 가독성을 위하여  $t$ 값 및 유의확률은 생략하였다.

Model 1a의 경우, 범죄발생은 시간에 따라 선형적 증가추세를 보여 해당기간동안 매월 약 67.36건 정도 강력범죄가 증가해왔을 알 수 있다. Model 1a와 Model 1b에서 고려하는 내재적 변동요인 중 시간 추세 다음으로 범죄발생에 영향을 끼치는 것은 전월 범죄발생건수 ( $y_{t-1}$ )와 전년도 해당월의 범죄발생건수 ( $y_{t-12}$ )임을 알 수 있다.

Model 2a와 Model 2b의 경우, 내재적 요인에 월별효과를 추가적으로 반영한 결과, 1월, 2월 가변수 회귀계수 값이 유의한 음 (-)의 값으로 나타난 것을 볼 수 있는데, 이는 1, 2월에 범죄 발생 건수가 저점을 보이는 반복적인 양상을 반영하였음을 알 수 있다. 한편 일반적으로 여름철 강력범죄가 높을 것이라는 예상과는 달리, 여름철 가변수는 유의하게 나타나지 않았는데, 이는 원계열 시도표상에서도 볼 수 있듯이, 최저점은 1,2월로 상당히 일정한 패턴으로 나타나고 있는데 비하여 고점 (peak)은 여름이 아닌 오히려 봄철 5월 혹은 가을철 9, 10월로 나타난 것에 기인한다. 월별 평균 강력범죄 발생건수를 보면, 1월의 평균범죄발생건수는 약 19,722건, 2월의 평균범죄발생건수는 약 21,104건으로 상당히 낮은 값을 보

이고 있는데 비해, 7월과 8월의 월별 평균 범죄발생건수는 각각 약 29,178건, 약 28,471건으로 12월의 평균발생건수(약 30,314)에 비해 유의한 차이를 보이지 않음을 확인할 수 있다. 물론 월별 평균범죄발생건수는 시간에 따른 증감을 고려하지 않은 단순 평균에 불과하므로 이를 근거로 해석하기에는 무리가 따를 수 있으나 원계열에 대한 시도표 증감패턴 및 이러한 월별 평균값을 총체적으로 판단하여 볼 때, 최근 10여년간 여름철 강력범죄 발생건수는 연말보다 유의하게 더 많지는 않은 것으로 판단된다. 오히려 5월의 경우 회귀계수 값이 유의한 양 (+)의 값을 보이며, 즉 5월은 다른 월에 비해 많은 강력범죄가 발생함을 나타내고 있는데, 이는 5월이 가정의 달, 가계 지출이 많은 달이라는 점에서 강력범죄 발생과의 연관성을 파악하고 이와 관련한 캠페인 등을 고려할 수 있음을 시사 한다고 할 수 있다.

Model 3a과 Model 3b는 계절성 순환주기를 고려하는 모형으로 확산계절변동은 유의하지 않았고, 고정계절 변동을 나타내는 삼각함수  $\sin(2\pi t/12)$ ,  $\cos(2\pi t/12)$ 이 통계적으로 유의하기는 하였으나, 모형의 간명성 (parsimony) 및 적합도 측면에서 볼 때, Model 1a와 Model 1b에 비해 효용이 없으며, 확산계절변동이 없으므로 해석에의 용이성(interpretability) 측면에서 볼 때 Model 2a과 Model 2b의 월별 가변수로도 추세 및 월별 계절성을 충분히 반영하는 것으로 판단된다.

한편 오차항의 자기상관은 Model 1b, Model 2b, Model 3b 모두 13차만 유의하였으며, 이는 월별 강력범죄발생은 전년도 전월에 의존함을 나타낸다. 한편 이들 6개 모형에 비해 ARIMA 모형은 모형의 적합도 (AIC, SBC)는 좋으나, 2013년 1월부터 12월까지의 예측값에 대한 예측력 (root mean square; RMSE)이 크게 떨어지는 것을 확인할 수 있었는데, 이는 ARIMA 모형이 빠른 변동을 갖는 시계열에 적합한 모형으로 시계열의 정상성을 가정하고 있는데 비해, 2013년 이후의 범죄발생 자료의 성격이 기존의 양상과는 다소 달라져 정상성 가정에 어긋났기 때문일 것으로 예상된다. 반면 시계열 회귀모형 및 자기회귀 오차항을 갖는 시계열 회귀모형은 ARIMA 모형 보다는 덜 민감하게 변동하는 보다 강건한 모형이므로 예측력이 더 나음을 알 수 있다.

**Table 4.1** Estimated coefficients of each model considering endogenous factors

| Variable            | Coefficient estimates |          |          |          |          |          |
|---------------------|-----------------------|----------|----------|----------|----------|----------|
|                     | Model 1a*             | Model 1b | Model 2a | Model 2b | Model 3a | Model 3b |
| Intercept           | 3945                  | 3377     | 7452     | 7624     | 5897     | 5295     |
| $t$                 | 67.36                 | 51.28    | 102.43   | 109.16   | 102.80   | 87.56    |
| $y_{t-1}$           | 0.38                  | 0.39     | 0.33     | 0.32     | 0.31     | 0.33     |
| $y_{t-3}$           |                       |          | 0.20     | 0.19     |          |          |
| $y_{t-12}$          | 0.35                  | 0.40     |          |          | 0.26     | 0.30     |
| Jan                 |                       |          | -8512    | -8298    |          |          |
| Feb                 |                       |          | -4164    | -4216    |          |          |
| March               |                       |          |          |          |          |          |
| April               |                       |          |          |          |          |          |
| May                 |                       |          | 3459     | 3663     |          |          |
| June                |                       |          |          |          |          |          |
| July                |                       |          |          |          |          |          |
| August              |                       |          |          |          |          |          |
| September           |                       |          |          |          |          |          |
| October             |                       |          |          |          |          |          |
| November            |                       |          |          |          |          |          |
| $\sin(2\pi t/12)$   |                       |          |          |          | -720.94  | -566.28  |
| $\cos(2\pi t/12)$   |                       |          |          |          | -1100.00 | -1037.00 |
| $t \sin(2\pi t/12)$ |                       |          |          |          |          |          |
| $t \cos(2\pi t/12)$ |                       |          |          |          |          |          |
| AR(13)              |                       | 0.24     |          | 0.27     |          | 0.26     |
| AIC                 | 2088.77               | 2083.11  | 2215.57  | 2207.46  | 2187.62  | 2180.67  |
| SBC                 | 2099.50               | 2096.52  | 2234.90  | 2229.56  | 2203.71  | 2199.45  |
| RMSE                | 3536.34               | 3545.81  | 3764.88  | 4024.61  | 3711.00  | 3563.00  |

\*Note that the number of each model in Table 4.1 corresponds to the number of formula in Table 3.1.



한편, Table 4.2는 외재적 요인으로 고려하는 변수들과 범죄발생의 관련성, 그리고 외재적 요인들 간의 다중공선성 문제를 확인하기 위하여 상관분석을 실시하고 각 변수에 해당하는 분산팽창계수 (variance inflation factor)를 확인한 결과이다. 상관분석 결과, 각 변수들과 5대 강력범죄 수의 연관성이 높아 이들을 외재적 요인으로 범죄발생을 설명하는 것이 타당함을 확인할 수 있다. 이들 발생 요인들 간에도 어느 정도 강한 상관관계를 보이고는 있으나, GDP와 경찰 1인당 담당인구 수를 제외하고는 일반적으로 다중공선성 문제가 우려되는 0.8 0.9 수준보다 하회 하였으며, GDP와 외국인 유입수와 경찰1인당 담당인구, 그리고 검거자 수에서 나타나는 강한 상관관계 역시 강한 추세에 기인한 것이므로 추세항을 포함하는 회귀 모형 내에 독립변수로서 함께 고려하기에 무리가 없을 것으로 판단하여 이들을 모두 포함하여 외재적 모형을 적합하였다. 분산팽창인자인 VIF 역시 모두 10이내의 값으로 다중공선성이 문제되지 않는 것으로 확인하였다 (Kutner 등. 2004).

**Table 4.2** Correlation coefficient of exogenous factors and variance inflation factor

|             | $y_t$    | $x_{t1}$ | $x_{t2}$ | $x_{t3}$ | $x_{t4}$ | $x_{t5}$ | $x_{t6}$ | VIF   |       |
|-------------|----------|----------|----------|----------|----------|----------|----------|-------|-------|
| crimes      | $y_t$    | 1        |          |          |          |          |          |       |       |
| gdp         | $x_{t1}$ | 0.885    | 1        |          |          |          |          | 8.227 |       |
| unemp       | $x_{t2}$ | 0.663    | 0.758    | 1        |          |          |          | 3.810 |       |
| foreigner   | $x_{t3}$ | 0.787    | 0.876    | 0.742    | 1        |          |          | 5.789 |       |
| r_male      | $x_{t4}$ | -0.683   | -0.712   | -0.652   | -0.515   | 1        |          | 3.460 |       |
| arrest      | $x_{t5}$ | 0.791    | 0.751    | 0.785    | 0.63     | -0.793   | 1        | 4.500 |       |
| npop_police | $x_{t6}$ | -0.81    | -0.889   | -0.822   | -0.818   | 0.749    | -0.821   | 1     | 7.712 |

Table 4.3은 내재적 요인과 외재적 요인을 모두 고려한 시계열 모형들 (Table 3.1의 (4)~(6)번 유형), 즉 시계열 회귀모형과 오차항의 자기상관을 갖는 회귀모형의 모형적합 결과이다. 추세에 외적 요인을 함께 고려한 Model 4a과 Model 4b는 경제적 요인으로는 실업자 수 (unemp), 사회변동요인으로는 16~24세 남성인구 비율 (r\_male)과 구조변화 (Date.Change) 및 교호효과 (Change\_rmale)가 강력범죄발생에 영향을 끼치는 유의한 외적 요인으로 나타났으며, 범죄 통제요인으로는 경찰 1인당 담당인구수 (npop\_police)가 강력범죄발생에 영향을 끼치는 유의한 외적 요인으로 나타났다. 이때 Date.Change는 16~24세 남성 인구비 시계열에서 2008년 6월부터 나타나는 구조변화를 반영하기 위한 변수로  $C_{t*} = \begin{cases} 1, & \text{since Jun2008,} \\ 0, & \text{otherwise.} \end{cases}$  와 같이 정의할 수 있으며, 이의 교호효과 (Change\_rmale)는  $r_{male} \times C_{t*}$ 를 나타낸다. 젊은 남성인구에 대한 구조변화의 교호작용이 유의한 것으로 보아, 2008년 6월 이전에는 젊은 남성인구가 줄어들지만 범죄발생이 늘어났던 것에 반해, 2008년 6월 이후에는 젊은 남성인구 증가가 강력범죄발생에 유의한 양(+)의 영향을 끼치고 있음을 알 수 있다.

한편, 외국인 유입수의 경우 지수적 증가 추세를 보이고 분산이 다소 커지는 양상을 보여 로그변환을 고려할 수 있으나, 오히려 기존에 유의하였던 GDP 등의 외재적 요인의 유의성이 없어지고 예측력 또한 매우 낮아져 원변수 그대로 모형 적합하였다. 비록 외국인 유입수의 유의성은 Model 4a부터 Model 6b 어느 모형에서도 유의성이 입증되지는 않았으나, 최근 외국인 노동력 유입의 증가 추세 등을 고려할 때 젊은 남성 인구 비율 등의 효과로 반영되었을 가능성도 존재할 수 있으며, 향후 이러한 외국인 유입의 증가 추세가 어느 정도 지속될 것으로 보이므로 이에 대한 검토는 의미가 있다고 할 수 있다.

추세와 월별효과를 고려하는 Model 5a과 Model 5b의 경우, 외재적 요인으로 GDP와 검거자수 (arrest)가 선택되는 것으로 보아, 범죄 발생을 설명하는 외재적 경제, 사회, 통제 요인들이 월별효과에 어느 정도 잠재적으로 반영되었을 것으로 예상된다. Model 6a과 6b에서 고려하는 순환요인 삼각함수는 내재적 모형에서와 마찬가지로 모형의 복잡성에 비해 효용이 없어, 적합도 및 해석적 측면에서 볼 때 Model 5a과 Model 5b보다 나은 모형이라 할 수 없다.

2012년까지의 범죄발생만을 이용한 모형 적합도 (AIC, SBC) 측면에서는 내적요인으로 추세만을 고려하는 Model 4a과 Model 4b가 더 우수하나, 2013년 범죄발생에 대한 예측력 (RMSE) 측면에서는

Model 5a과 Model 5b가 더 우수함을 알 수 있다. 즉 강력범죄발생을 설명 및 예측하는데 가장 우수한 모형은 범죄발생의 내재적 변동요인인 추세 및 월별효과를 고려하는 한편, 외재적 요인 중 사회전반적인 경제상황을 나타내는 GDP, 범죄통제요인인 검거지수를 고려하는 모형이라 할 수 있다.

**Table 4.3** Estimated coefficients of each model with endogenous and exogenous factors

| Variable            | Coefficient estimates |          |          |          |          |          |
|---------------------|-----------------------|----------|----------|----------|----------|----------|
|                     | Model 4a*             | Model 4b | Model 5a | Model 5b | Model 6a | Model 6b |
| Intercept           | 239350                | 277761   | -3922    | -12647   | 47922    | 51291    |
| $t$                 | 80.06                 | 66.26    | 59.51    | 38.78    |          |          |
| $y_{t-1}$           | 0.17                  | 0.13     | 0.28     | 0.18     | 0.21     | 0.20     |
| $y_{t-3}$           | 0.18                  | 0.09     |          |          | 0.02     | -0.08    |
| $y_{t-12}$          | 0.17                  | 0.21     |          |          |          |          |
| Jan                 |                       |          | -8073    | -7361    |          |          |
| Feb                 |                       |          | -3745    | -3036    |          |          |
| March               |                       |          |          |          |          |          |
| April               |                       |          |          |          |          |          |
| May                 |                       |          |          |          |          |          |
| June                |                       |          |          |          |          |          |
| July                |                       |          |          |          |          |          |
| August              |                       |          |          |          |          |          |
| September           |                       |          |          |          |          |          |
| November            |                       |          |          |          |          |          |
| $\sin(2\pi t/12)$   |                       |          |          |          | -1457    | -1670    |
| $\cos(2\pi t/12)$   |                       |          |          |          | -1757    | -1552    |
| $t \sin(2\pi t/12)$ |                       |          |          |          |          |          |
| $t \cos(2\pi t/12)$ |                       |          |          |          |          |          |
| gdp                 |                       |          | 0.05     | 0.10     |          |          |
| unemp               | -5.74                 | -5.29    |          |          |          |          |
| r_male              | -824779               | -1041621 |          |          | -566008  | -607314  |
| foreigner           |                       |          |          |          |          |          |
| arrest              |                       |          | 0.03     | 0.03     | 0.04     | 0.05     |
| npop_police         | -181.41               | -236.18  |          |          |          |          |
| Date_Change         | -143503               | -200888  |          |          | -273505  | -323412  |
| Change_rmale        | 2330657               | 3241805  |          |          | 4398793  | 5191845  |
| AR(11)              |                       | 0.30     |          | 0.21     |          |          |
| AR(13)              |                       | 0.31     |          | 0.31     |          | 0.24     |
| AIC                 | 2073.16               | 2043.63  | 2254.09  | 2232.81  | 2264.81  | 2257.63  |
| SBC                 | 2099.98               | 2075.81  | 2276.32  | 2260.60  | 2289.67  | 2285.25  |
| RMSE                | 4657.29               | 5556.56  | 2519.03  | 2985.02  | 3725.00  | 3587.00  |

\*Note that the number of each model in Table 4.3 corresponds to the number of formula in Table 3.1.

#### 4.3. 모형의 평가

본 연구에서 고려하는 13개의 시계열 모형 중, ARIMA 모형을 제외한 나머지 내재적 모형 및 외재적 모형 중 적합도가 가장 우수한 모형은 범죄발생의 내재적 변동 중 추세를 반영하는 모형들 즉 Model 1a, Model 1b, Model 4a, Model 4b이었다. 특히 범죄발생의 내재적 증가 추세와 여러 외재적 요인, 그리고 젊은 남성인구의 구조변화를 함께 고려한 자기 회귀오차를 갖는 시계열 회귀모형 Model 4b가 가장 적합도가 우수하였고 그 다음으로는 Model 1b, Model 4a, Model 1a순이었다.

그러나 향후 1년간 (2013년 1월~2013년 12월)의 예측력이 가장 우수한 모형은, 내재적 요인으로는 추세 외에 월별효과까지 고려하는 동시에 외재적 요인들을 함께 고려한 모형 Model 5a, Model 5b이 가장 우수하였다. 특히 과거 범죄 발생정도와 계절효과 그리고 범죄발생에 대한 여러 위험요인 및 이의 구조적 변화까지 고려한 시계열 회귀모형인 Model 5a가 가장 우수하였으며 그 다음으로는 Model 5b, Model 6a의 예측력이 좋았다.

이처럼 모형을 평가함에 있어서 적합도와 예측력 측면에서 우수한 모형이 동일한 것은 아니다. 이는 적합도는 기존의 자료를 가장 잘 설명할수록 우수한 반면, 예측력은 새로운 자료를 잘 설명할수록 우수하기 때문인데, 기존의 자료에 너무 의존하는 모형은 새로운 자료가 기존의 양상과 약간의 차이가 있을 때 이를 제대로 잘 반영하지 못하기 때문이다. 이는 동일 변수를 적합시키는 모형에 오차항에 대한 가정만을 다르게 하더라도 확인할 수 있는데, Table 4.1과 Table 4.3에서 오차항에 대해 자기상관을 고려하지 않은 모형과 자기상관을 고려하는 모형간에 적합도 및 예측력 차이를 보더라도 알 수 있다. 대체로 기존 자료에 대한 양상을 최대한 반영시키는 자기회귀오차항을 갖는 회귀모형이 적합도는 더 좋으나, 반대로 예측력은 떨어지는 것을 볼 수 있는데, 오차항에 대한 상관을 반영함으로써 기존의 자료에 최적화된 과적합 (overfitted) 때문으로 예상된다. 따라서 모형의 적합도와 예측력 측면에서 우수한 모형은, 강력 범죄발생의 내재적 증가추세 및 변동을 고려하는 동시에 범죄발생에 영향을 끼칠 수 있는 외재적 요인까지 반영하는 Model 5a이 가장 강건 (robust)하게 우수함을 알 수 있다.

Table 4.4는 예측력이 가장 좋은 Model 5a, Model 5b, Model 6a과 Model 6b의 범죄 발생 건수의 실제값과 각 모형의 예측값을 정리한 표이다. 예측 결과, Model 6a과 Model 6b이 Model 5a과 Model 5b에 비해 예측력이 떨어지는 이유는 범죄발생 예측에 있어서 겨울의 과대예측 및 여름의 과소예측 정도가 크기 때문이다. 가장 예측력이 좋은 Model 5a 역시 이와 같은 겨울철 범죄발생에 과대예측, 여름철 범죄발생에 과소예측을 보이고 있는데, 이는 Model 5a에서 유의한 월별효과가 11, 1, 2월 즉 겨울철 범죄발생 효과만 반영된 데 비해, 2013년에는 이전보다 여름에 강력범죄가 더 많이 발생하고, 다른 계절에는 강력범죄 발생이 줄어들었기 때문으로 보인다.

Table 4.4 Predicted number of violent crime outbreaks

| Time    | Observed value | Predicted value |          |          |          |
|---------|----------------|-----------------|----------|----------|----------|
|         |                | Model 5a*       | Model 5b | Model 6a | Model 6b |
| 2013.01 | 29,985         | 31888.78        | 31821.89 | 38193.01 | 36723.33 |
| 2013.02 | 28,589         | 32209.87        | 34528.57 | 35697.22 | 38263.95 |
| 2013.03 | 33,105         | 35951.79        | 37754.17 | 36289.46 | 37471.53 |
| 2013.04 | 35,389         | 38093.36        | 38454.16 | 38206.01 | 39450.82 |
| 2013.05 | 37,629         | 39433.30        | 40583.82 | 39976.04 | 41801.08 |
| 2013.06 | 42,290         | 40169.43        | 38834.28 | 41642.73 | 42018.21 |
| 2013.07 | 42,230         | 41300.67        | 40093.36 | 43293.79 | 44257.03 |
| 2013.08 | 40,835         | 41365.37        | 39939.79 | 43347.31 | 43691.10 |
| 2013.09 | 38,359         | 41627.80        | 39755.07 | 42588.99 | 42819.22 |
| 2013.10 | 42,172         | 41753.28        | 42648.07 | 41109.31 | 42219.89 |
| 2013.11 | 35,673         | 40594.94        | 38678.11 | 40813.31 | 40068.87 |
| 2013.12 | 41,003         | 40903.49        | 41127.84 | 38533.92 | 39233.88 |

\*Note that the number of each model in Table 4.4 corresponds to the number of formula in Table 3.1.

## 5. 결론

본 연구에서는 강력범죄 발생의 요인을 시간, 계절, 순환의 내재적 요인과 경제, 사회변동 및 범죄통제와 관련된 외재적 요인으로 나누고 다양한 시계열 회귀모형 및 오차항의 자기상관을 고려한 시계열 회귀모형을 적합하고 그 적합도와 예측력을 비교하였다. 내재적 요인 변수와 외재적 요인 변수들을 함께 고려함에 있어 전이함수 모형을 고려하였으나 다중 전이함수로의 적합이 가능하지 않아 본 논문에서는 포함하지 않았다.

시계열 회귀 모형과 자기상관 시계열 회귀모형의 회귀계수들의 적합 결과 내재적 요인이 외재적 요인에 비하여 그 영향력이 큼을 확인할 수 있었으며 범죄발생의 내재적 변동 요인 중 추세와 계절성은 시간 변수, 시차변수, 그리고 월별효과에 대한 가변수로 충분히 반영할 수 있음을 확인할 수 있었다. 범죄발생의 내재적 변동요인 외에 범죄발생에 영향을 끼치는 외재적 요인까지 모두 고려하는 모형이 적합도와

예측력을 개선함을 확인할 수 있었으며, 가장 적합도와 예측력이 우수한 모형을 기준으로 볼 때, 범죄발생에 영향을 끼치는 외재적 요인으로는 전반적인 경제상황을 반영하는 GDP와 범죄통제상황을 반영하는 검거자수가 가장 유의함을 알 수 있었다. 즉 GDP가 낮아지는 등 전반적인 경제상황이 어려워지면 강력범죄가 발생할 가능성이 높는데, 검거와 같은 강력한 범죄 통제를 통해 강력범죄 발생을 낮추거나 예방할 수 있음을 시사한다. 또한 12월을 기준으로 볼 때, 11, 1, 2월의 범죄발생건수가 낮은 것으로 보아, 겨울보다는 봄부터 가을까지, 특히 여름철 강력범죄발생 예방을 위한 적절한 범죄 통제 정책이 필요함을 시사한다고 할 수 있다. 향후 범죄발생과 관련하여 검거자수 이외에 범죄발생에 영향을 끼치는 기타 통제요인 등의 공공자료가 추가적으로 수집 및 공표된다면, 이를 모형에 반영함으로써 범죄발생 예측 시 나타나는 예측오류를 더 개선할 수 있을 것으로 예상된다.

## References

- Beki, C., Zeelenberg, K. and Montfort, K. (1999). An analysis of the crime rate in the Netherlands 1950-93. *British Journal of Criminology*, **39**, 401-415.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, **28**, 591-605.
- Choi, Y. (2007). Illegal stay of foreigners and their crime. *Korean Criminological Review*, **18**, 1319-1340.
- Corman, H. and Mocan, H. N. (2000). A time-series analysis of crime, deterrence, and drug abuse in New York city. *The American Economic Review*, **90**, 584-604.
- Corman, H., Joyce, T. and Lovitch, N. (1987). Crime, deterrence and the business cycle in New York city: A VAR approach. *The Review of Economics and Statistics*, **69**, 695-700.
- Deadma, D. and Pyle, D. (1997). Forecasting recorded property crime using a time-series econometric model. *British Journal of Criminology*, **37**, 437-445.
- Godfrey, L. (1978). Testing for high order serial correlation in regression equations when the regressors include lagged endogenous variables. *Econometrica*, **46**, 1303-1310.
- Ju, I. and Cho, K. (2012). Analysis of seasonal factors on crime occurrence. *Journal of Korean Public Police and Security Studies*, **8**, 101-124.
- Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2004). *Applied linear regression models*, 4th Ed., McGraw-Hill Irwin, Chicago.
- Kim, H. (2004). *The effect of the green space on the crime in the city*, Master Thesis, Hongik University, Seoul.
- Lee, H. (2015). Analysis of statistical models on temperature at the Suwon city in Korea. *Journal of the Korean Data & Information Science Society*, **26**, 1409-1416.
- Lee, S. (2007). A study on the foreigners' crimes in Korea. *Korean Police Studies Review*, **6**, 79-110.
- Lee, T. (2010). *A study on crime factors : Focusing on urban areas*, Master Thesis, Kyunghee University, Seoul.
- Lim, S. S. (2014). A study on the forecasting models using housing price index. *Journal of the Korean Data & Information Science Society*, **25**, 65-76.
- Park, M. R. and Kim, Y. S. (2015). Crime risk implementation for safe return service. *Journal of the Korean Data & Information Science Society*, **26**, 1097-1104 .
- Park, S., Jang, A. and Lee, J. (2009). Social change and crime trends in Korea, 1966-2007. *Social Research*, **18**, 45-72.
- Park, Y. and Kim, K. (2002). *Time series analysis using SAS/ETS 1*, Freedom Academy, Seoul.
- Park, Y. and Song, S. (1998). *Business & economic data analysis using SAS/ETS*, Freedom Academy, Seoul.
- Park, H. (2002). *A study on the relationship between crime rates and economic indicator: focusing on unemployment*, Master Thesis, Korea University, Seoul.
- Ryu, S. R. and Kim, J. T. (2013). Time series regression model for forecasting the number of elementary school teachers. *Journal of the Korean Data & Information Science Society*, **24**, 321-332.
- Shah Habibullah, M. and Baharom, A. H. (2009). Crime and economic conditions in Malaysia. *International Journal of Social Economics*, **36**, 1071-1081.

## A study on the violent crime and control factors in Korea<sup>†</sup>

Tae Yeon Kwon<sup>1</sup> · Saebom Jeon<sup>2</sup>

<sup>1</sup>Institute for Economic Research, Korea University

<sup>2</sup>Department of Statistics, Korea University

Received 21 September 2016, revised 31 October 2016, accepted 5 November 2016

### Abstract

The increasing trend of the five violent crimes (murder, robbery, rape, violence, theft) in Korea is not independent of social and economic factors. Several social science research have discussed about this issue but most of them do not properly reflect the nature of the time-series data. Based on several time series models, we studied about the endogenous factors (time, seasonal and cycle factors) and exogenous factors (economical, social change and crime control factors) on violent crime occur in Korea. Autocorrelation were also taken into account. Through this study, we want to help to make preventive policy by explaining the cause of violent crime and predicting the future incidence of it.

*Keywords:* Autocorrelation, crimes outbreaks, endogenous factors, exogenous factors, time series regression model.

---

<sup>†</sup> This work is partially supported by a Korea University Grant.

<sup>1</sup> Research professor, Institute for Economic Research, Korea University, Seoul 02841, Korea.

<sup>2</sup> Corresponding author: Research professor, Department of Statistics, Korea University, Seoul 02841, Korea. E-mail: [alwaysns@korea.ac.kr](mailto:alwaysns@korea.ac.kr)