

## 단일변수 시계열 모형들의 주택가격지수 예측력 비교

### Comparing the Forecasting Performance of Univariate Time Series Models with Korean House Price Index

이 영 수 (Lee, Youngsoo)\*

#### < Abstract >

This paper compares the forecasting performance of four types of univariate time series models: ARIMA, GARCH, regime-switching, and unobserved-component(trend-cycle decomposition) model. I carry out in-sample and out-of-sample forecasting exercise to Korean house price index in the recursive estimation scheme. For the comparison of forecast performances, I compute root mean squared forecast errors and mean absolute forecast errors for each forecasting horizon of 1 to 36 months. I also test the statistical significance of the forecast comparison results, applying the asymptotic test developed by Diebold and Mariano(1995).

In case of in-sample forecasts, regime-switching and unobserved-component models show better performance than other two models, along with statistical significance. Regime-switching model is a compelling choice until 12 month forecasting steps, while the unobserved-component model is the best for 24 and 36 months. In case of out-of-sample forecasts, I find that the regime-switching model is particularly ill-suited for forecasting, showing the same result as Crawford and Fratantoni(2003) with US house price data. Also, I find that the unobserved-component model shows the best out-of-sample forecasting performance.

주 제 어 : 주택가격, 예측, ARIMA, IGARCH, 국면전환, 비관측요인 모형

key word : House Price, Forecast,, ARIMA, IGARCH, Regime-switching, Unobserved-component

\* 영산대학교 부동산금융학과 교수, yslee@ysu.ac.kr

## I. 서론

전 세계를 휩쓸었던 ‘서브 프라임 모기지 사태’ 이후, 주택가격 전망의 중요성이 더욱 크게 부각되고 있다. 사실, 주택가격에 대한 높은 관심은 여러 경제 참여자들에게서 두로 나타나는 현상이다. 가계자산에서 주택이 차지하는 비중이 높은 상황에서,<sup>1)</sup> 주택 소유자나 투자자로서의 일반 가계는 주택가격의 향배에 큰 관심을 갖게 되며, 주택가격과 경기와의 관련성이 높다는 점에서 경제 정책 담당자들의 관심도 높을 수밖에 없다.<sup>2)</sup> 또한 주택담보 대출 비중이 높은 금융기관 역시 주택가격 변화에 대해 항상 주의를 기울이게 된다.<sup>3)</sup> 이에 따라 주택가격에 대한 보다 정확한 예측은 개인투자자, 경제정책 당국, 금융기관 등 여러 경제 주체들 모두에게 주요한 관심사가 되고 있다.

주택가격 예측에서의 효시적 논문은 Case and Shiller(1989)이다. Case and Shiller는 미국 주택시장에서의 정보 효율성을 검토한 결과, 주택가격이 임의보행(random walk) 과정을 따르지 않는 것을 발견하였다.<sup>4)</sup> 주택가격이 임의보행을 하지

않는다면, 여러 가지 시계열 모형을 이용하여 주택가격에 대한 예측이 가능할 수 있음을 의미한다.

Case and Shiller 이후 시계열 모형을 이용한 주택가격 예측 모형의 개발이 활발히 이루어져 왔다. 기본적인 ARIMA에서부터 VAR, 거시계량 모형 등을 이용한 전망을 비롯하여, 최근에는 GARCH 모형, 국면전환 모형, 동적 요인 모형(dynamic factor model)등이 활용되면서 주택가격 예측 모형에 대한 논의가 활발히 진행되고 있다.

본 연구에서는 여러 가지 시계열 모형 중 단일 변수 시계열 모형들을 정리하고, 이러한 모형들의 예측력을 우리나라의 주택가격지수(아파트가격지수)를 이용하여 비교하고자 한다. 단일 변수 모형이 갖는 한계점에도 불구하고, 주택가격과 관련이 있는 여러 경제 변수들을 적기에(timely) 획득하기 어렵다는 점 등을 고려할 때 주택가격 자체 변수만을 이용하는 단일 변수 시계열 모형은 나름대로의 의의를 갖고 있다고 생각된다.

본 연구에서와 같이 단일 변수 시계열 모형을 이용한 대표적인 논문으로는 Crawford and Fratantoni(2003)가 있다. 이들은 ARIMA, GARCH, 국면전환모형을 이용하여 미국의 지역별 주택가격에 대한 예측력을 비교하였다. 본 연구에서는

- 1) 통계청의 가계자산조사(2006년)에 의하면, 우리나라의 경우 가계자산에서 차지하는 주택 자산의 비중이 39.9%이다.
- 2) IMF에서의 연구결과에 의하면, 1970년에서 2000년 사이에 OECD 국가들에서 20번의 주택경기침체가 발생하였는데 그중 19번은 경기침체로 이어졌다고 발표한 바 있다. (Mills(2008, p250)) 참조
- 3) Crawford and Rosenblatt(1995)은 주택가격의 변동성이 주택대출 채무불이행(default)과 조기대출상환(prepayment)을 결정하는 주요한 요인이라는 분석결과를 발표하였다.
- 4) 이러한 주택시장의 비효율성이 나타나는 이유로 Malpezzi(1999)는 주택시장이 현금화하기 어려운 시장이라는 점, 주택들이 매우 이질적이며 투자대상이 아닌 소비재로 인식되는 경우가 많다는 점, 세금을 포함한 거래비용 및 주택에 관한 정보 획득 비용이 높다는 점 등을 들고 있다. 이러한 요인들은 우리나라의 경우에도 동일하게 적용될 수 있을 것이며, 손경환 외(2004), 이용만(2012), 이영수·박상현(2013) 등은 우리나라의 주택시장이 효율적이지 않음을 실증적으로 밝히고 있다.

위의 세 가지 모형 외에 비관측요인 모형(추세-순환 분해모형)도 추가하여 네 가지 단일변수 시계열 모형의 예측력을 비교하고자 한다.<sup>5)</sup>

본 연구에서는 Crawford and Fratantoni(2003)와는 달리 축차적 추정(recursive estimation)을 이용한 동적 예측(dynamic forecasting)치를 이용하였다. 모형의 비교를 위해 예측시계별(forecasting horizon; k step forecasting)로 RMSFE(root mean squared forecast error)와 MAFE(mean absolute forecast error)를 계산하였으며, 모형간의 예측력 차이가 통계적 유의성을 갖고 있는가에 대한 검정도 실시하였다.

본 연구의 구성은 다음과 같다. I 장의 서론에 이어, II 장에서는 선행연구를 검토하였다. III 장에서는 네 가지 단일변수 모형에 대한 간략한 설명과 함께 모형의 추정 결과 및 그 의미를 살펴보았으며, IV 장에서는 모형의 예측 결과들을 정리하였다. V 장은 결론이다.

## II. 선행연구 검토

시계열 예측 모형은 크게 단일 변수 시계열 모형과 다중 시계열 모형으로 구분된다. 단일 변수 시계열 모형으로는 본 연구에서 다루는 ARIMA, GARCH, 국면전환모형, 비관측요인모형 등이 있으며, 다중 시계열 모형을 이용한 주택가격 예측은 VAR, VECM(vector error correction model), BVAR(bayesian VAR), FAVAR(factor augmented VAR), DSGE(dynamic stochastic general

equilibrium), DFM(dynamic factor model)등을 통해 이루어진다. 여기서는 단일 시계열 모형을 중심으로 국내외 선행연구를 검토하고자 한다.<sup>6)</sup>

단일 시계열 모형을 이용한 간단한 모형은 ARIMA 모형이다. 이 모형의 특징은 주택가격 변동의 동학(dynamics)을 선형으로 가정한다는 점이다. Muellbauer and Murphy(1997)는 여타 경제변수와의 상호연관성, 높은 거래비용 등으로 이러한 선형성이 깨어질 수 있으며, 주택가격 변동의 동학은 시간에 따라 달라지는 가변성을 갖고 있다는 점을 영국의 주택가격 데이터를 통해 실증적으로 보여주었다. 이러한 분석 결과는 ARIMA 모형을 벗어난 여타 예측 시계열 모형의 필요성을 강조하고 있다.

주택가격을 예측한 논문 중 가장 많이 회자되는 논문 중의 하나는 Crawford and Fratantoni(2003)이다. Crawford and Fratantoni는 ARIMA, GARCH, 국면전환모형을 이용하여 미국의 지역별 주택가격에 대한 예측력을 비교하였는데, 논문의 초점은 국면전환모형의 예측력이 ARIMA 모형에 비해 높은가의 여부이었다. 앞서 Hall et al.(1997), Spiegel(2001) 등은 실증분석을 통해 주택시장이 호황기와 침체기로 구분되는 국면을 갖고 있음을 보여준 바 있다. Crawford and Fratantoni는 이러한 주택시장의 국면을 감안한 국면전환모형이 높은 예측력을 가질 수 있는가를 실증적으로 검토하고자 하였다. 그러나 실증분석 결과는 표본내(in-sample) 적합성(fitness)의 경우에는 국면전환모형이 우수하나 표본외(out-of-sample) 예측에서는 간단한 ARIMA 모형이 오히려 높은

5) 필자가 아는 한, 비관측요인모형에 의한 주택가격 예측은 국내외에서 아직 시도된 적이 없으며, GARCH 모형과 국면전환모형에 의한 주택가격 예측도 국내에서는 처음으로 시도되었다.

6) 다중 시계열 모형을 이용한 주택가격 예측과 관련된 논문들의 검토는 Gupta et al.(2011) 참조.

예측력을 갖고 있는 것으로 나타났다. ARIMA 모형과 GARCH 모형의 비교에서는 지역별·예측기간별로 다른 결과가 나타나고 있다.

Mills(2008)는 AR 모형에 비선형성을 감안하기 위하여 GAR(generalized AR) 모형을 이용하였다. Crawford and Fratantoni와 동일한 데이터를 이용한 분석에서, GAR 모형이 ARIMA, GARCH, 국면전환모형보다 예측력이 우수하다는 실증분석결과를 제시하였다.

주택가격 전망 모형과 관련된 또 하나의 흐름은 칼만 필터를 이용하는 방식이며, 주로 시변계수(time varying coefficient) 모형에 적용되고 있다. Brown et al.(1997)은 영국의 주택가격 데이터를 이용한 분석에서 시변계수 모형이 고정계수 모형보다 우수한 예측력을 갖고 있음을 발표하였으며, Guirguis et al.(2005)은 미국의 주택가격 데이터를 이용하여, 시변계수를 갖는 AR 모형이 통상적인 AR 모형이나 GARCH 모형보다 예측력이 우수하다는 분석 결과를 보고하였다.

우리나라의 주택가격을 이용한 시계열 예측 논문을 정리하면 다음과 같다.<sup>8)</sup> 김근용(1998)은 주택가격지수와 전세가격지수의 데이터를 이용하여 ARIMA 모형과 상태공간모형<sup>9)</sup>의 예측력을 비교하였다. 주택가격지수의 경우는 ARIMA 모

형이 전세가격지수에 대해서는 상태공간모형이 높은 예측력을 갖는 것으로 나타났다. 다만 이러한 결과는 1988년 1월부터 1996년 12월까지의 기간에 대한 추정 결과를 바탕으로 1997년 1월부터 9월까지의 단 9개월간의 예측결과만을 이용하고 있다는 점에서 두 모형의 비교 결과를 일반화시키는 데에는 무리가 따르는 것으로 생각된다.

윤주현·김혜승(2000)은 ARIMA 모형과 개입모형(intervention model)을 이용하여 주택가격에 대한 예측력을 검토하였으며, 개입모형이 ARIMA 모형에 비해 예측력이 다소 높은 것으로 보고하고 있다. 손정식 외(2002)는 토지가격, 주택매매가격, 전세가격에 대해 ARIMA 모형과 VAR 모형의 예측력을 비교하였다. 실증 분석결과는 VAR 모형의 예측력이 다소 높은 것으로 보고하였다.

박천규·이영(2010)은 여러 가지 주택시장 체감지표들의 주택가격 예측력을 검토하였는데, 매매수요 BSI<sup>10)</sup>가 가장 유용하다는 결과를 제시하였다. 한편, 황영진(2013)은 동태적 모형 선택법<sup>11)</sup>의 주택가격 예측력을 분석하였다. 비교대상 모형은 임의보행모형, AR 모형, 시변계수 AR 모형이다. 분석 결과는 동태적 모형 선택법의 예측력이 가장 우수하며, 시변계수 AR 모형이 통상

7) GAR 모형은 AR 모형에서 과거 변수들이 선형의 연결 구조를 갖는 것과는 달리, 과거 변수들이 비선형으로 연결되는 구조를 갖는다. (예컨대,  $y_t = a_1y_{t-1} + a_2y_{t-2} + a_3y_{t-3}^2 + a_4y_{t-1}y_{t-2} + \varepsilon_t$ )

8) 1990년대에 이루어진 거시계량모형이나 VAR 분석을 통한 다변량 예측모형에 대한 논의는 손정식 외(2002) 참조.

9) 김근용(1998)의 상태-공간 모형은 본 연구의 비관측요인 모형과 기본적인 틀은 유사하나, 모형의 형태는 크게 차이가 나고 있다. 김근용의 상태공간 모형은 추세-순환 분해 모형에서 순환요인이 제거된 모형이다.

10) 매매수요 BSI는 매수세/매도세의 체감 동향을 이용하여 국민은행이 계산하여 발표하는 지수(매수우위 지수)이다.

11) 동태적 모형 선택법(dynamic model selection)은 예측에 유용할 것으로 예상되는 변수들로 예측 후보 변수군을 구성한 뒤, 매 시점마다 예측에 가장 유용한 변수들을 선택하여 예측을 진행하는 방법이다.

적인 AR 모형보다 높은 예측력을 갖고 있는 것으로 나타났다. AR 모형의 예측력이 임의보행 모형의 예측력보다 오히려 약하다는 결과가 흥미롭다.

### III. 모형의 설정 및 추정

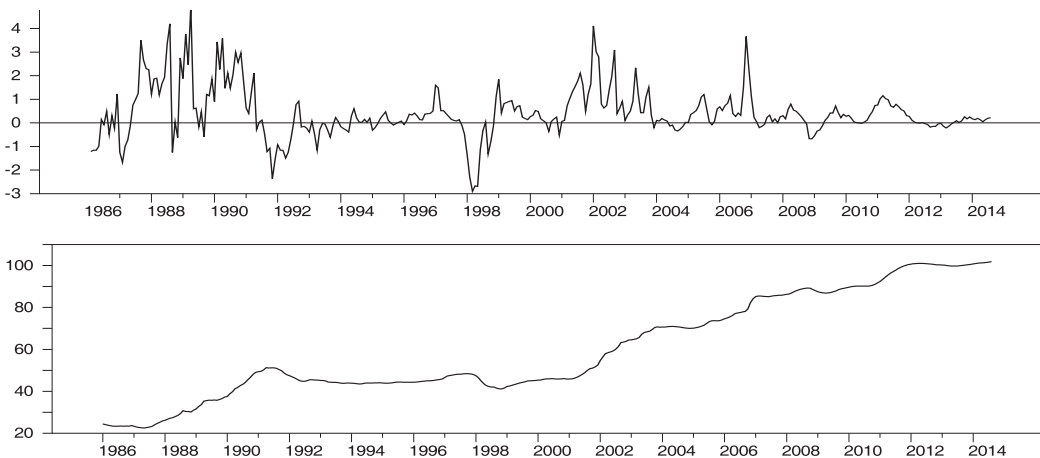
#### 1. 데이터 및 모형설정 개관

본 연구에서는 주택가격 변수로 국민은행이 발표하는 전국아파트가격지수(계절조정치)를 사용하였다. 국민은행의 전국주택가격지수를 이용할 수도 있으나, 주택에서 차지하는 아파트의 비중이 높고, 주택가격지수안에 포함되는 단독주택의 경우 주택 간의 이질성이 매우 높으며 거래량도 적기 때문에 지수 자체에 대한 신뢰성 문제가 발생할 수 있다는 점 등을 고려하였다. 변수는

로그로 변환하여 사용하였다. 전체 데이터 기간은 1986년 1월부터 2014년 8월까지이다. <그림 1>은 아파트가격지수와 그 증가율의 추이를 보여 준다.

본 연구에서는 네 가지 단일변수 시계열 모형을 설정하고, 모형들의 예측력을 비교한다. 첫 번째 모형은 시계열 모형 중 가장 기본적인 ARIMA 모형이다. 두 번째 모형은 GARCH 모형이다. <그림 1>에서 보는 것처럼 주택가격은 급격히 변동하기도 하고 일정기간 정체되기도 한다. Engle(1982), Bollerslev(1986) 등에 의해 개발된 ARCH/GARCH 모형은 이러한 변동성 집중의 특성을 갖는 시계열을 모형화하기 위해 사용된다.<sup>12)</sup> 세 번째 모형은 국면전환모형이다. 경기변동이 상승과 하강 등의 국면을 갖고 있는 것처럼 주택가격 변동도 국면에 종속되어 발생할 가능성이 있다. 국면전환 모형은 이러한 구조 변화를 담아내기 위한 모형이다. 네 번째 모형은 비관측

<그림 1> 아파트가격지수와 증가율



주: 아래 그림은 아파트가격지수(2013년 3월=100)이며, 위 그림은 지수의 전월대비 증가율이다.

12) 우리나라 주택가격의 경우에도 ARCH/GARCH 효과가 존재하고 있는 것으로 분석되고 있다. (임재만, 2006; 한용석 외, 2010)

요인 모형(추세-순환분해)이다. 주택가격을 변동시키는 외부 충격은 다양한 성격을 가질 것이다. 추세-순환 분해 모형에서는 이러한 충격을 영구적인 효과를 갖는 충격과 일시적인 효과를 갖는 충격으로 구분한다.

모형 설정에서 고려되어야 할 사항중의 하나는 구조변화이다. 주택시장은 대표적인 불완전시장의 하나로서 정부의 정책에 크게 영향을 받으며, 외부효과가 매우 크게 작용하는 시장이다. 이러한 특징으로 인해 주택가격 결정 체계에 구조적 변화가 일어날 공산이 높다. 구조변화 더미변수, 개입모형, 시변계수(파라미터) 모형 등은 이러한 구조적 변화를 감안한다. 선행연구 검토에서 살펴본 것처럼 구조변화를 감안한 모형의 예측력도 높아지는 효과가 있다. 그러나 본 연구에서는 이러한 구조변화를 감안하지 않은 기본적인 형태로 모형을 설정하였다. 모형 비교를 위해서는 구조적 변화를 네 가지 모형 모두에서 고려할 필요가 있는데, 모형의 설정이나 추정상 어려운 점이 있다는 현실적 제약이 한 요인이며, 다른 한편으로는 구조변화를 감안하지 않는데 따른

비효율성이 네 가지 모형 모두에서 발생할 것이기 때문에 모형 비교에는 큰 영향이 없을 것이라는 점도 고려하였다.

## 2. ARIMA 모형

단일변수 시계열모형 중 예측을 위해 널리 사용되는 모형의 하나는 ARIMA 모형이다. ARIMA(p,1,q) 모형은 다음과 같이 표현된다.

$$\Delta y_t = \mu(1 - \sum_{i=1}^p \alpha_i) + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \sum_{i=0}^q \beta_i \varepsilon_{t-i} \quad (1)$$

$$\varepsilon_t \sim N(0, \sigma^2) \quad (2)$$

위에서  $\Delta$ 는 1차 차분을 의미하며,<sup>13)</sup>  $\alpha_i$ 는 AR의 계수 그리고  $\beta_i$ 는 MA의 계수이다.

본 연구에서는 ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(2,1,1), ARIMA(2,1,2)의 4가지 모형을 적용하였다. ARIMA(0,1,0)는 확률보행(random walk) 모형으로서 일반적인 ARIMA 모형에 의한 예측

〈표 1〉 ARIMA 모형의 추정결과

	ARIMA(0,1,0)	ARIMA(1,1,0)	ARIMA(2,1,1)	ARIMA(2,1,2)
$\mu$	0.416*** (0.056)	0.432*** (0.143)	0.444*** (0.158)	0.436*** (0.148)
$\alpha_1$	-	0.727*** (0.036)	-0.173*** (0.045)	-0.199*** (0.052)
$\alpha_2$	-	-	0.745*** (0.037)	0.694*** (0.053)
$\beta_1$	-	-	0.848*** (0.051)	0.930*** (0.072)
$\beta_2$	-	-	-	0.117* (0.071)
log L	-504.771	-372.292	-350.556	-349.288

주: ( )안은 표준오차이며, \*, \*\*, \*\*\*는 각각 10%, 5%, 1%의 수준에서 계수가 유의적임을 표시한다.

13) 아파트가격지수는 Augmented Dickey-Fuller 및 Phillipse-Perron 단위근 검정 결과, 1% 유의수준에서 1차 차분 안정적(first difference stationary)인 것으로 검정되었다.

과의 차이를 비교하는 기준을 제공해준다는 점을, ARIMA(1,1,0)은 가장 간단한 형태의 ARIMA 모형이라는 점을, 그리고 ARIMA(2,1,1)과 ARIMA(2,1,2)는 각각 BIC 기준과 AIC 기준에 따른 적정모형이라는 점을 감안하였다.<sup>14)</sup> 모형의 추정은 최우추정법을 사용하였으며, <표 1>은 추정결과이다.

### 3. GARCH

ARCH/GARCH 모형의 예비 추정결과, 조건부 분산과정이 불안정적(unstationary)이어서 본 연구에서는 IGARCH(1,1) 모형을 선택하였다. 평균방정식을 ARIMA 형태로 설정한 IGARCH 모형은 다음과 같다.

$$\Delta y_t = \theta + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \sum_{i=0}^q \beta_i \varepsilon_{t-i} \quad (3)$$

$$\varepsilon_t | \Omega_{t-1} \sim N(0, \sigma_t^2) \quad (4)$$

$$\sigma_t^2 = c + a * \varepsilon_{t-1}^2 + b * \sigma_{t-1}^2 \quad (5)$$

식 (3)은 ARIMA 과정을 나타내는 방정식이며,  $\beta_0=1$ 이다. 식 (4)는 오차항의 조건부 분산이 식 (2)와는 달리 시간가변적임을 의미하며, 식 (5)는 그러한 오차항의 조건부 분산 과정을 GARCH (1,1)으로 표현하고 있다. IGARCH는  $a+b =1$ 의 제약을 갖는다.

본 연구에서 평균방정식은 ARIMA(1,1,0), ARIMA(2,1,1), ARIMA(2,1,2)의 세 가지를 이용하였으며, <표 2>는 최우추정법에 의한 추정결과이다. 표에서 보는 바와 같이 GARCH의 계수인 a와 b는 모두 높은 통계적 유의성을 갖고 있어서, 우리나라의 주택가격지수가 ARCH/GARCH 효과를 갖고 있음을 보여주고 있다.

### 4. 국면전환모형

국면전환모형(RS; regime-switching model) 은 시계열변수의 확률과정이 비관측상태변수(국면)

<표 2> ARIMA-IGARCH 모형 추정결과

	ARIMA(1,1,0)-IG	ARIMA(2,1,1)-IG	ARIMA(2,1,2)-IG
$\theta$	0.019 (0.014)	0.033 (0.025)	0.046 (0.034)
$\alpha_1$	0.793*** (0.031)	-0.079 (0.076)	-0.115 (0.107)
$\alpha_2$	-	0.714*** (0.065)	0.658*** (0.089)
$\beta_1$	-	0.839*** (0.085)	0.972*** (0.132)
$\beta_2$	-	-	0.168*** (0.074)
c	0.008*** (0.002)	0.007** (0.002)	0.005** (0.002)
a	0.365*** (0.060)	0.341*** (0.058)	0.300** (0.049)
b	0.634*** (0.060)	0.658*** (0.058)	0.699*** (0.049)
log L	-225,226	-222,988	-220,827

주: ( )안은 표준오차이며, \*, \*\*, \*\*\*는 각각 10%, 5%, 1%의 수준에서 계수가 유의적임을 표시한다.

14) 적정모형 선정을 위한 최대 차수(lag)는 AR과 MA에서 모두 4로 제한하였다.

에 종속되는 모형이다. 국면에 따라 변동되는 파라미터를 어디까지 어떻게 설정하는가에 따라 여러 가지 형태가 있을 수 있으나, 본 연구에서는 Hamilton(1989)의 마코프 국면전환 모형을 이용한다. 모형의 형태는 다음과 같다.

$$\Delta y_t - \mu(S_t) = \sum_{i=1}^p \alpha_i * (\Delta y_{t-i} - \mu(S_{t-i})) + \varepsilon_t \quad (6)$$

위에서  $S_t$ 는 국면을 표시하며, 본 연구에서는  $S_t$ 가 1과 2의 두 가지 국면을 갖는 것으로 상정한다. 이 경우,  $S_t=1$ 은 주택가격의 증가율이 낮은 침체기, 그리고  $S_t=2$ 는 주택가격의 증가율이 높은 상승기로 해석될 수 있다. 식 (6)은 상태종속 평균으로부터의 이탈치( $\Delta y_t - \mu(S_t)$ )가 AR(p) 과정을 따르는 것을 의미한다.

국면  $S_t$ 는 마코프 특성<sup>15)</sup>을 갖는 것으로 설정하며, 다음과 같은 전이행렬(transition matrix)을 갖는다.

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} \quad (7)$$

여기서  $p_{ij} = \Pr(S_{t+1}=j | S_t=i)$ 로서, 국면이  $i$ 에서 다음 기에  $j$ 로 전이될 확률을 표시한다. 2국면 모형에서  $p_{11}+p_{12}=1$ ,  $p_{21}+p_{22}=1$ 이다.

본 연구의 국면전환모형에서  $y_t$ 의 확률과정은 ARIMA(1,1,0)와 ARIMA(2,1,0)의 두 가지 모형을 고려하였다. 모형의 추정은 베이즈 정리를 적용한 최우추정법을 이용하였으며, <표 3>은 추정

결과이다.

<표 3> 국면전환모형의 추정결과

	ARIMA(1,1,0)-RS	ARIMA(2,1,0)-RS
$\mu(st=1)$	0.327*** (0.143)	0.329*** (0.144)
$\mu(st=2)$	2.379*** (0.205)	2.394*** (0.206)
$\alpha_1$	0.806*** (0.034)	0.767*** (0.057)
$\alpha_2$	-	0.046 (0.056)
p11	0.959*** (0.011)	0.960*** (0.011)
p21	0.673*** (0.148)	0.674*** (0.150)
log L	-313.007	-312.762

주: ( )안은 표준오차이며, \*, \*\*, \*\*\*는 각각 10%, 5%, 1%의 수준에서 계수가 유의적임을 표시한다.

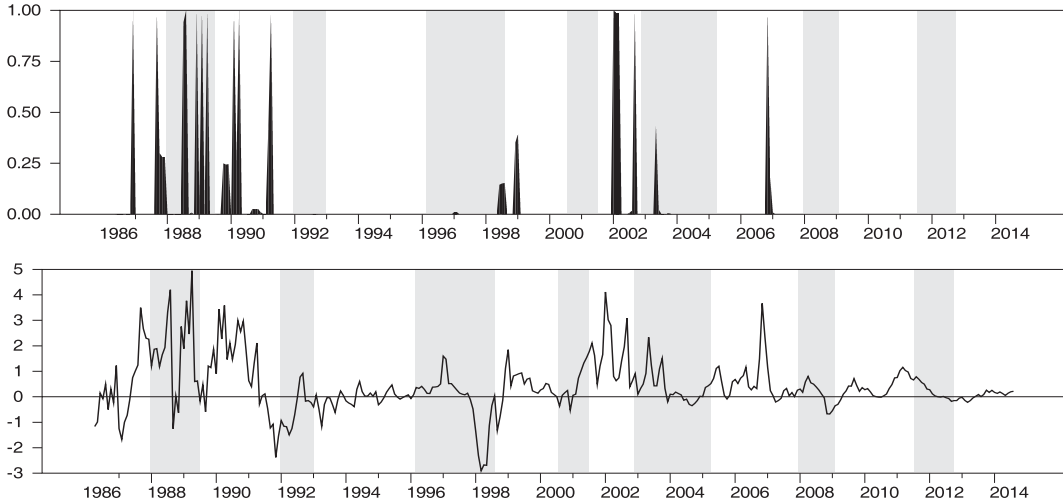
<표 3>에서 보는 바와 같이  $\alpha_2$ 는 통계적 유의성을 갖지 못하며, ARIMA(1,1,0)-RS 모형이 보다 통계적으로 적절한 모형인 것으로 나타나고 있다. ARAIMA(1,1,0)-RS 모형의 추정결과를 통해 우리나라 아파트가격 변동의 특징을 살펴보면 다음과 같다. 첫째, 아파트가격 상승국면에서의 평균가격 증가율은 2.38% 그리고 침체국면에서의 평균가격 증가율은 0.33%이다. 둘째, 상승국면보다는 침체국면이 일반적이다. 위의  $p_{ij}$ 를 이용하여 국면이 상승국면 혹은 침체국면에 있을 비조건부 확률을 구하면,<sup>16)</sup> 국면이 침체국면에 있을 비조건부 확률은 94.4%, 상승국면에 있을 비조건부 확률은 5.6%로서, 아파트가격 국면이 침체국면에 있을 경우가 압도적이다. 셋째, 국면이 침체국면에 있는 경우가 압도적이라면, 이는 침체국면으로 해석하는 것보다는 통상적인 국면으로 해석하는 좋을 것이다. 이러한 결과에 입각하여 유추하면, 아파트가격이 급락하는 경우보다는 아파트가격이 급등하는 경우가 보다 일반적

15) 마코프 특성은 국면  $S_t$ 의 변화 확률이 일정하며, 그 확률값은 전기의 상태국면( $S_{t-1}$ )에만 의존하는 특성을 의미한다.

16) 이러한 확률을 ergodic 확률이라고 하며,  $\Pr(S_t=1) = (1-p_{22}) / (2-p_{11}-p_{22})$ 의 공식을 이용하여 계산할 수 있다.



〈그림 2〉 상승국면에 대한 평활화 확률 (ARIMA(1,1,0)-RS 모형)



주: 위의 그림은 주택가격국면이 상승국면에 있을 확률에 대한 평활화 확률 그림이며, 아래는 아파트가격지수 증가율이다. 음영부분은 통계청에서 발표한 경기하강국면을 표시한다.

라는 결론에 다다른다. 이는 아파트가격 변동 국면을 둘로 나누는 경우, 통상적인 국면과 급락국면이 아니라 통상적인 국면과 급등국면으로 구분되기 때문이다.

<그림 2>는 ARIMA(1,1,0)-RS 모형에 의거한 상승국면의 평활화 확률- $\Pr(S_t=1|\Omega_T)$ -을 보여준다.  $\Pr(S_t=1|\Omega_T)$ 는 전기간에 걸친 데이터( $\Omega_T$ )를 이용하여 계산한,  $t$ 기에서의 국면이 상승(급등)국면일 확률이다. 그 확률이 0.5를 넘어서는 경우는 많지 않으며, 그러한 상승국면이 지속될 확률(17)도 낮아서 검은 색 막대가 매우 좁게 표시되고 있다. 주택가격 상승(급등) 국면은 80년대 말과 90년대 초, 2002년에 두 차례, 2006년 말에 한차례 발생하였다. 80년대 말의 상승 국면은 경기 하강국면 속에서 나타났는데, 1기 신도시 건설붐과 관련이

있는 것으로 생각된다. 여타 상승(급등) 국면은 모두 경기상승기의 중반에 나타나고 있다.

### 5. 비관측요인 모형

다른 시계열 변수들과 마찬가지로 주택가격 역시 추세(trend)와 순환(cycle) 요인으로 분해할 수 있다. 본 연구에서는 비관측요인(unobserved component; UC) 모형을 이용하여 추세-순환 분해를 시도한다.<sup>18)</sup> 모형의 형태는 다음과 같다.

$$y_t = \tau_t + c_t \tag{8}$$

$$\tau_t = \beta + \tau_{t-1} + u_t, \quad u_t \sim \text{i.i.d. } N(0, \sigma_u^2) \tag{9}$$

17) 상승국면이 지속될 확률( $p_{22}$ )은  $1-p_{21}$ 이며, 추정치를 이용한 계산값은 32.7%이다.

18) 비관측요인 모형을 이용한 추세-순환 분해 논의는 주로 GDP를 중심으로 이루어져 왔다. 주택가격에 대한 추세-순환 분해 논문으로는 미국의 데이터를 이용한 Clark and Coggin(2009), 한국의 데이터를 이용한 이영수·박상현(2013) 등이 있다.

$$c_t = \phi_1 c_{t-1} + \phi_2 c_{t-2} + v_t, v_t \sim \text{i.i.d. } N(0, \sigma_v^2) \quad (10)$$

위에서  $\tau_t$ 와  $c_t$ 는 각각 추세치와 순환치를 의미하며, 추세는 식 (9)의 확률보행과정으로 순환은 식 (10)의 AR(2) 과정으로 표현되고 있다.  $u_t$ 와  $v_t$ 는 오차항들이다. 비관측요인모형의 효시적 논문인 Harvey(1985)와 Clark(1987)에서는 두 오차항의 독립성을 가정하였으나, Morley et al.(2003)은 그러한 제약이 없는 모형을 제시하였다. 본 연구에서는 두 가지 경우를 모두 고려한다.

두 오차항( $u_t$ 와  $v_t$ )의 상관계수는 1보다 크거나 -1보다 작을 수 없다. 이러한 제약을 모형의 추정에 반영하기 위해, 식 (10)의  $v_t$ 를 다음과 같이 변환한다.

$$v_t = a * u_t + e_t, \quad e_t \sim \text{i.i.d. } N(0, \sigma_e^2) \quad (11)$$

위에서  $u_t$ 와  $e_t$ 는 독립적이다. 계수  $a$ 는  $v_t$ 와  $u_t$ 의 상관성의 정도를 나타내는 계수로서  $a=0$ 이면,  $v_t$ 와  $u_t$ 의 상관계수는 0이 되며,  $a>0$ 이면 +의 상관계수 그리고  $a<0$ 이면 -의 상관계수가 될 것이다. 또한  $e_t$ 의 분산이 0인 경우 상관계수는 +1 혹은 -1이 될 것이다. 이제 식 (11)을 이용하여  $u_t$ 와  $v_t$ 의 상관계수( $\rho_{uv}$ )를 구하면 다음과 같다.

$$\rho_{uv} = a * \sigma_u / (a^2 \sigma_u^2 + \sigma_e^2)^{1/2} \quad (12)$$

식 (8)-(11)로 구성되는 비관측요인모형은 관찰되지 않는 변수를 포함하고 있어 직접적인 추정이 불가능하다. 모형을 추정하기 위한 상태-공간

모형으로 재구성하면 다음과 같다.

$$y_t = (1 \ 1 \ 0) \begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix} \quad (13)$$

$$\begin{pmatrix} \tau_t \\ c_t \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \phi_1 & \phi_2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \tau_{t-1} \\ c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} \beta \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ a & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_t \\ e_t \end{pmatrix} \quad (14)$$

위에서 식 (13)은 관측방정식으로 불리며, 관측치  $y_t$ 와 상태벡터(state vector)를 연결시키는 식이다. 식 (14)는 상태방정식으로 상태벡터의 전이 과정을 나타낸다. 이러한 상태-공간 모형의 추정은 칼만 필터(Kalman filter)를 이용한 최우추정법을 사용한다. <표 4>는 추정결과이다.

<표 4>의  $\sigma_u$ ,  $\sigma_e$ ,  $a$ 의 추정치를 식 (12)에 대입하면<sup>19)</sup> UC-UR 모형에서 추세 오차항( $v_t$ )과 순환 오차항( $u_t$ )의 상관계수를 계산할 수 있다. 계산 결과는 -1.0에 거의 일치하는 값이다. 상관계수가 -1이라면, Morley et al.(2003)이 지적한 것처럼 ARIMA 모형과 구조상 차이를 갖지 않게 된

<표 4> 비관측요인모형의 추정결과

	UC-0	UC-UR
$\beta_0$	0.398*** (0.055)	0.406*** (0.129)
$\varphi_1$	1.804*** (0.040)	0.714*** (0.053)
$\varphi_2$	-0.808***(0.040)	-0.064***(0.027)
$\sigma_u$	0.321*** (0.059)	2.465*** (0.335)
$\sigma_e$	0.588*** (0.050)	0.000 (0.411)
$a$	-	-0.713*** (0.037)
log L	-370.185	-367.841

주: ( )안은 표준오차이며, \*, \*\*, \*\*\*는 각각 10%, 5%, 1%의 수준에서 계수가 유의적임을 표시한다.

19)  $\sigma_e$ 의 추정치는 0.00000103이다.

다. 따라서 여기서는 UC-0의 모형을 이용하여, 비관측요인 모형이 시사하는 아파트가격변동의 특징을 살펴본다.

첫째, 아파트 가격 변동은 추세요인이 아니라 주로 순환요인에 의해 발생한다는 점이다. <그림 3>에서 보는 바와 같이 추세선은 매우 평탄하며, 주택가격의 변동은 주로 순환요인에 의해 나타나고 있다. 둘째, 이러한 아파트가격 순환에는 일정한 주기가 존재하지 않는다. 식 (10)과 같은 AR(2) 방정식에서 순환요인  $c_t$ 가 일정한 주기를 갖기 위해서는 일정한 필요조건을 만족시켜야 하는데,<sup>20)</sup> 추정된  $\phi_1$ 과  $\phi_2$ 는 이러한 필요조건을 만족시키지 못하고 있다. 셋째, 한번 발생한 아파트 가격 변동은 <그림 3>에서도 쉽게 확인할 수 있는 것처럼 가속성과 함께 지속성을 갖는다.<sup>21)</sup> 즉 아파트가격의 상승-하락 사이클의 기간이 매우 길게 나타나고 있다. 넷째, 1980년대 후반을 제외하고는 경기하강기에 아파트 가격도 하강 사이

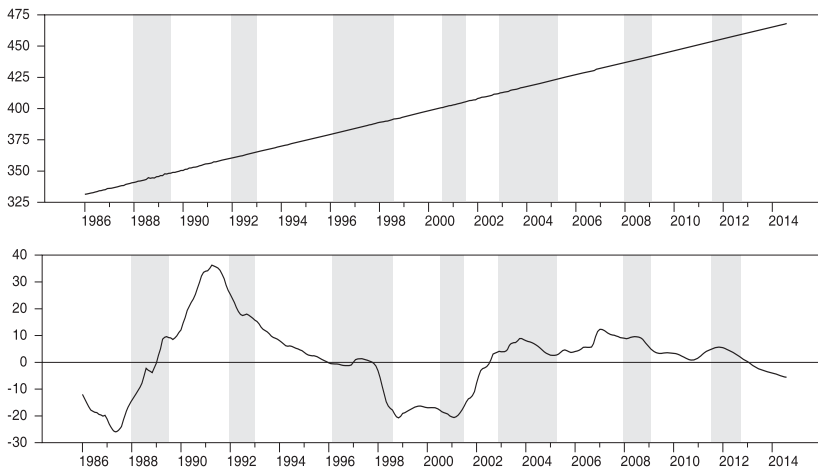
클을 보이고 있어, 경기와의 관련성이 존재하고 있음을 시사하고 있다. 다섯째, 최근의 아파트가격은 추세요인 밑에서 움직이고 있으며, 아파트 가격지수의 순환요인의 방향은 하락 사이클을 보이고 있다.

## IV. 모형의 예측력 비교

### 1. 모형별 예측력

본 연구에서는 표본내(in-sample) 예측과 표본외(out-of-sample) 예측을 모두 실시한다. 표본내 예측은 데이터 전체 기간(1986년 1월~2014년 8월)에 대한 모형의 추정 결과를 이용하며, 표본외 예측은 축차적 추정(recursive estimation)의 결과를 이용한다. 축차적 추정은 1986년 1월부터

<그림 3> 아파트가격지수의 추세-순환 분해 (UC-0 모형)



주: 위의 그림은 추세, 아래 그림은 순환이며, 음영부분은 통계청에서 발표한 경기하강국면을 표시한다.

20) 식 (10)의 후방연산자 다항식  $B(L)$ 이 허수근을 가져야 한다.

21) 이러한 특성은 식 (10)에 대한 충격반응함수를 통해 확인된다.

1999년 12월까지의 기간을 시작으로 하여, 1개월씩 추정기간을 늘려가며 시행하였다. 표본내 예측기간은 1986년 4월부터 2014년 8월까지, 그리고 표본외 예측기간은 2000년 1월부터 2014년 8월까지이다.

예측은 수준변수( $y_t$ )에 대해 이루어졌으며, 예측오차는 다음과 같이 실측치에 대한 비율로 계산하였다.

$$\varepsilon_{t+i} = \{(\hat{y}_{t+i} - y_{t+i}) / y_{t+i}\} * 100 \quad (15)$$

위 식에서  $\hat{y}_{t+i}$ 는 예측시계(forecasting horizon)  $i$ 에 대한  $t$ 기에서의 동적예측치(dynamic forecast)

이다.

모형의 예측력 비교를 위해 각 예측시계별로 평균제곱예측오차의 제곱근(root mean squared forecast error; 이하 RMSFE)과 평균절대오차(mean absolute forecast error; 이하 MAFE)를 계산하였다.

$$RMSFE(i) = \left( \sum_{t=1}^n \varepsilon_{t+i}^2 / n \right)^{0.5} \quad (16)$$

$$MAFE(i) = \sum_{t=1}^n |\varepsilon_{t+i}| / n \quad (17)$$

위에서  $i$ 는 예측시계 그리고  $n$ 은 예측치의 수

〈표 5〉 예측오차: ARIMA 모형

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	ARIMA(0,1,0)	0.282	0.521	0.747	1.335	2.329	3.768	4.834
	ARIMA(1,1,0)	0.197	0.367	0.574	1.095	2.083	3.602	4.708
	ARIMA(2,1,1)	0.184	0.363	<b>0.564</b>	<b>1.080</b>	<b>2.050</b>	<b>3.577</b>	4.692
	ARIMA(2,1,2)	<b>0.183</b>	<b>0.363</b>	0.565	1.089	2.070	3.593	<b>4.683</b>
		(MAFE)						
	ARIMA(0,1,0)	0.186	0.354	0.518	0.967	1.738	2.848	3.664
	ARIMA(1,1,0)	0.113	0.230	0.375	0.754	1.516	2.728	3.581
ARIMA(2,1,1)	<b>0.109</b>	<b>0.229</b>	<b>0.369</b>	<b>0.742</b>	<b>1.486</b>	<b>2.715</b>	3.587	
ARIMA(2,1,2)	0.109	0.230	0.371	0.751	1.502	2.721	<b>3.563</b>	
표 본 외		(RMSFE)						
	ARIMA(0,1,0)	0.175	0.330	0.469	0.826	1.478	2.467	2.822
	ARIMA(1,1,0)	<b>0.116</b>	<b>0.245</b>	<b>0.378</b>	0.691	1.333	2.359	2.784
	ARIMA(2,1,1)	0.117	0.247	0.383	<b>0.690</b>	<b>1.318</b>	2.344	2.799
	ARIMA(2,1,2)	0.116	0.245	0.381	0.690	1.319	<b>2.341</b>	<b>2.777</b>
		(MAFE)						
	ARIMA(0,1,0)	0.116	0.228	0.333	0.621	1.119	1.841	<b>2.023</b>
	ARIMA(1,1,0)	<b>0.068</b>	<b>0.155</b>	<b>0.251</b>	0.510	1.008	1.775	2.053
ARIMA(2,1,1)	0.071	0.156	0.255	<b>0.507</b>	1.007	1.772	2.111	
ARIMA(2,1,2)	0.070	0.156	0.254	0.509	<b>1.005</b>	<b>1.759</b>	2.069	

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다.

이다.

<표 5>는 ARIMA 모형들에 대한 예측력을 보여주고 있다. 그 결과는 다음과 같다. 첫째, ARIMA 모형은 확률보행 모형-ARIMA(0,1,0)-에 비해 높은 예측력을 갖는다. 이러한 결과는 표본내 예측만이 아니라 표본외 예측에서도 나타난다. 다만, 이러한 높은 예측력은 예측시계가 장기로 갈수록 약해진다. 둘째, 표본내 예측에서 단순한 ARIMA(1,1,0)보다는 ARIMA(2,1,1)이나 ARIMA(2,1,2)가 높은 예측력을 갖고 있다. 표본외 예측에서는 예측시계 3개월까지의 단기에서는 ARIMA(1,1,0)의 예측력이 다소 높으나, 그 이후의 예측시계에서는 ARIMA(2,1,1)이나 ARIMA(2,1,2)에 비해 예측력이 떨어진다. 셋째, ARIMA(2,1,1)과 ARIMA(2,1,2)를 비교하면, 표본 내에서는 ARIMA(2,1,1)이 우수하며, 표본 외에서는

ARIMA(2,1,2)가 다소 높은 예측력을 보이고 있다.

<표 6>은 IGARCH 모형들의 예측력 결과이다. 표본내 예측과 표본외 예측 모두에서, ARIMA(2,1,2)-IGARCH의 예측력이 가장 약하며, RMSFE 기준으로는 ARIMA(2,1,1)-IGARCH 모형이 그리고 MAFE 기준으로는 ARIMA(1,1,0) -IGARCH 모형이 예측력 면에서 상대적으로 우수한 것으로 나타나고 있다.

국면전환모형의 경우, 식 (6)에서 국면  $S_t$ 는 관측이 불가능하며 따라서 국면에 종속되는 계수값  $\mu(S_t)$ 를 알 수 없다는 문제점이 있다. 본 연구에서는 국면이 발생할 확률로 계산한 기대치를 사용하였다.

$$\mu^e = \mu(s_t=1)*P_t(S_t=1) + \mu(s_t=2)*P_t(S_t=2) \quad (18)$$

<표 6> 예측오차: IGARCH 모형

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	ARIMA(1,1,0)-IG	0.198	0.369	0.577	1.088	2.052	3.578	4.692
	ARIMA(2,1,1)-IG	0.189	<b>0.366</b>	<b>0.571</b>	<b>1.084</b>	<b>2.043</b>	<b>3.572</b>	4.690
	ARIMA(2,1,2)-IG	<b>0.189</b>	0.367	0.571	1.091	2.061	3.584	4.677
		(MAFE)						
	ARIMA(1,1,0)-IG	0.113	<b>0.227</b>	<b>0.369</b>	<b>0.736</b>	1.481	<b>2.703</b>	3.580
	ARIMA(2,1,1)-IG	<b>0.111</b>	0.229	0.371	0.740	<b>1.480</b>	2.708	3.595
	ARIMA(2,1,2)-IG	0.113	0.232	0.375	0.752	1.499	2.718	<b>3.569</b>
	표 본 외		(RMSFE)					
ARIMA(1,1,0)-IG		<b>0.116</b>	0.245	<b>0.378</b>	0.688	1.324	2.351	2.788
ARIMA(2,1,1)-IG		0.117	0.246	0.381	<b>0.687</b>	<b>1.316</b>	<b>2.338</b>	<b>2.788</b>
ARIMA(2,1,2)-IG		0.116	<b>0.244</b>	0.379	0.691	1.330	2.361	2.794
		(MAFE)						
ARIMA(1,1,0)-IG		<b>0.067</b>	<b>0.153</b>	<b>0.248</b>	<b>0.501</b>	<b>0.995</b>	1.769	<b>2.069</b>
ARIMA(2,1,1)-IG		0.072	0.156	0.256	0.507	1.003	<b>1.767</b>	2.106
ARIMA(2,1,2)-IG		0.075	0.158	0.259	0.515	1.010	1.779	2.082

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다. IG는 IGARCH를 의미한다.

〈표 7〉 예측오차: 국면전환(RS) 모형

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	ARIMA(1,1,0)-RS	0.127	0.262	0.421	0.893	1,707	2,743	3.499
	ARIMA(2,1,0)-RS	0.127	0.260	0.418	0.889	1,712	2,784	3.562
		(MAFE)						
	ARIMA(1,1,0)-RS	0.084	0.181	0.295	0.627	1,223	2,096	2,673
	ARIMA(2,1,0)-RS	0.085	0.179	0.291	0.622	1,207	2,106	2,711
표 본 외		(RMSFE)						
	ARIMA(1,1,0)-RS	0.115	0.243	0.375	0.686	1,326	2,345	2,770
	ARIMA(2,1,0)-RS	0.120	0.255	0.394	0.753	1,489	2,688	3,239
		(MAFE)						
	ARIMA(1,1,0)-RS	0.066	0.151	0.245	0.498	0,988	1,767	2,045
	ARIMA(2,1,0)-RS	0.068	0.153	0.244	0.520	1,051	1,940	2,227

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다.

위에서  $P_t(S_t=j)$ 는  $t$ 시점에  $j$ 국면이 발생할 확률이다. 국면전환모형의 추정과정에서 이러한 확률에 대한 추정치를 구할 수 있으며, 이 추정치를 이용하여  $\mu_t^e$ 를 구하고 그 값을 이용하여 표본내 예측오차를 계산할 수 있다.

한편, 표본외 예측의 경우에도, 예측기간중의  $\mu_{t+h}^e$ 를 어떻게 계산할 것인가의 문제가 발생한다. 이 경우에도 앞서와 마찬가지로 국면이 발생할 확률을 감안한 예측치를 사용하였다. 식 (7)의 전이행렬과 식 (18)을 이용하면 다음과 같이  $\mu_{t+h}^e$ 에

〈표 8〉 예측오차: 비관측요인(UC) 모형

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	UC-0	0.173	0.328	0.515	0.982	1,789	2,810	3,365
	UC-UR	0.192	0.369	0.578	1.110	2,088	3,543	4,631
		(MAFE)						
	UC-0	0.098	0.203	0.329	0.664	1,252	1,986	2,402
	UC-UR	0.111	0.232	0.377	0.764	1,504	2,631	3,466
표 본 외		(RMSFE)						
	UC-0	0.119	0.250	0.381	0.670	1,255	2,118	2,455
	UC-UR	0.119	0.250	0.387	0.704	1,351	2,409	2,855
		(MAFE)						
	UC-0	0.068	0.156	0.250	0.499	0,974	1,563	1,801
	UC-UR	0.069	0.159	0.257	0.518	1,008	1,782	2,018

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다.

대한 예측치를 계산할 수 있다.

$$\hat{\mu}_{t+i}^e = [P_t(S_t=1) \ P_t(S_t=2)] \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}^{-1} \begin{bmatrix} \mu(S_t = 1) \\ \mu(S_t = 2) \end{bmatrix} \quad (19)$$

<표 7>은 위의 과정들을 통해 계산된 국면전환모형들의 예측력 결과이다. 그 결과들을 정리하면 다음과 같다. 첫째, 표본내 예측에서는 6개월까지의 단기의 경우 ARIMA(2,1,0)에 국면전환모형을 적용한 경우가 예측력이 상대적으로 높았으며, 2년 이후의 장기에는 ARIMA(1,1,0)에 국면전환모형을 적용한 경우가 예측력이 상대적으로 높았다. 둘째, 표본외 예측의 경우, ARIMA(1,1,0)-RS 모형이 ARIMA(2,1,0)-RS 모형보다 높은 예측력을 보이고 있다.

<표 8>은 비관측요인모형들의 예측력 결과이다. 추세와 순환간의 상관성을 감안한 UC-UR 모형에 비해 두 요인간의 관련성을 감안하지 않는 UC-0 모형이 예측력에서 크게 앞서는 것으로 나타났다.

## 2. 예측력 비교

본 절에서는 모형간의 예측력을 비교한다. 예측력의 비교는 기준모형의 예측력에 대한 상대적인 비교로써 이루어지며, 이러한 상대적 차이가 통계적 유의성을 갖고 있는가에 대한 검정도 시행한다. 예측력 차이 검정은 Diebold and

Mariano(1995)에 의해 제안된 점근적 검정(asymptotic test) 방법을 이용하되,<sup>22)</sup> 소규모 표본 편이(bias)를 감안한 Harvey et al.(1997)의 수정방식을 적용하였다.

<표 9>는 ARIMA(1,1,0) 관련 모형들의 예측력 비교이다. 기준 모형은 ARIMA(1,1,0)이며, 비교 모형은 ARIMA(1,1,0)-IGARCH와 ARIMA(1,1,0)-RS 모형이다. 표에서의 수치는 비교 모형의 예측오차(RMSFE와 MAFE) 값들을 기준 모형의 값으로 나눈 수치로서, 그 수치가 1보다 작으면 예측력이 기준 모형에 비해 우수하다는 것을 의미한다.

표에서 보는 바와 같이 표본내 예측의 경우, 국면전환모형(ARIMA(1,1,0)-RS)의 예측력이 가장 높은 것으로 나타나고 있다. 예측시계에 상관없이 국면전환모형의 예측오차의 RMSFE와 MAFE 값들은 기준 모형인 ARIMA(1,1,0) 모형에 비해 80% 이하에 불과하며, 대부분 1%~10%의 유의수준에서 그 차이가 유의적인 것으로 검정되고 있다. 그러나 표본외 예측의 경우에는 그 결과가 크게 달라지고 있다. 표본내 예측에서와는 반대로 국면전환모형의 예측력이 가장 낮으며, 이러한 낮은 예측력은 대부분 통계적 유의성을 갖는 것으로 나타나고 있다.<sup>23)</sup> 예측시계가 길어질수록 예측오차 비율이 점차 커지면서 ARIMA 모형에 비해 예측력이 더욱 낮아지는 것도 특기할 만하다.

국면전환모형의 표본내 예측력이 높은 것은 주택가격의 평균증가를 계수가 ARIMA 모형이나

22) Diebold and Mariano 검정은 예측오차의 손실함수값(본 연구에서는 squared error 혹은 absolute error)의 차(loss-differential)의 분포가 점근적으로 정규분포한다는 이론적 근거 하에 이루어진다.

23) 국면전환모형에 의한 예측이 표본내 예측에서는 우수하나 표본외 예측에서는 오히려 타 모형에 비해 예측력이 크게 떨어지는 현상은 미국 주택가격을 이용한 Crawford and Fratantoni(2003)의 분석 결과와 일치하고 있다.

<표 9> 예측력 비교: ARIMA(1,1,0) 관련 모형

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	ARIMA(1,1,0)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	ARIMA(1,1,0)-IG	1.001	1.006	1.012	1.002	0.993	1.000	1.012**
	ARIMA(1,1,0)-RS	0.656**	0.692**	0.699**	0.763**	0.765*	0.755*	0.771
		(MAFE)						
	ARIMA(1,1,0)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ARIMA(1,1,0)-IG	0.988	0.990	1.001	0.988	1.000	0.998	1.030***	
ARIMA(1,1,0)-RS	0.782***	0.781***	0.768***	0.782**	0.806**	0.798*	0.763**	
표 본 외		(RMSFE)						
	ARIMA(1,1,0)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	ARIMA(1,1,0)-IG	0.998	0.999	1.001	0.996	0.993	0.996	1.001
	ARIMA(1,1,0)-RS	1.055***	1.078***	1.095***	1.121**	1.121*	1.119	1.135
		(MAFE)						
	ARIMA(1,1,0)	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ARIMA(1,1,0)-IG	0.983***	0.984	0.987*	0.983*	0.987	0.996	1.007	
ARIMA(1,1,0)-RS	1.084***	1.081***	1.079**	1.097**	1.104*	1.097*	1.129**	

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다. IG는 IGARCH를 의미한다.

IGARCH 모형에서처럼 고정되어 있지 않고 상태 국면에 따라 변동하며, 표본내 예측에서 이러한 변동에 대한 기대치를 사용하였기 때문일 것이다.<sup>24)</sup> 그러나 표본외 예측에서는 이러한 변동 기대치의 사용이 예측력을 오히려 크게 약화시키고 있는 점이 주목된다. Bessec and Bouabdallah (2005)는 이러한 현상이 미래 국면에 대한 오분류(misclassification of future regimes)에 기인하고 있다고 설명하고 있다.

한편, ARIMA 모형과 IGARCH 모형을 비교하면, 표본내 예측력에서는 예측시계 36개월을 제외하고는 별 차이가 없는 것으로 나타나고 있다. 예측시계 36개월의 경우, ARIMA 모형의 예측력

이 통계적으로 유의하게 높은 것으로 나타났다. 표본외 예측에서는 IGARCH 모형이 예측시계 36개월을 제외하고는 예측력이 다소 높다.

<표 10>은 예측력이 상대적으로 우수한 세부 모형들을 중심으로 각 모형간의 상대적 예측력을 비교하고 있다. ARIMA 모형에서는 ARIMA(2,1,2), IGARCH 모형에서는 ARIMA(2,1,1)-IGARCH(1,1), 국면전환 모형에서는 ARIMA(1,1,0)-RS 모형, 그리고 비관측요인 모형에서는 UC-0 모형을 이용하였다. 기본 모형은 비관측요인(UC-0) 모형이며, 상대적 예측력 계산 방식은 앞의 <표 9>에서와 같다.

표본내 예측의 경우, 12개월까지의 단기에서는

24) 식 (6) 및 (18) 참조.



국면전환 모형이 가장 높은 예측력을 보이고 있다. 국면전환 모형(ARIMA(1,1,0)-RS)은 앞서 살펴본 ARIMA(1,1,0) 관련 모형뿐만 아니라, 여타 ARIMA 관련 모형 그리고 비관측요인 모형보다 높은 예측력을 갖고 있다.

그러나 24개월과 36개월의 장기에서는 비관측요인 모형의 예측력이 국면전환 모형보다 높은 예측력을 보이고 있다. 다만, 그 차이에 대한 통계적 유의성은 검증되지 않는다. 또한 비관측요인 모형의 예측력은 ARIMA 모형이나 IGARCH 모형에 비해 모든 예측시계에서 높게 나타나며, 통계적 유의성

도 갖고 있는 것으로 검증된다. 비관측요인 모형의 예측 우수성이 예측시계가 길어질수록 상대적으로 높아지는 것도 특기할 만하다.<sup>25)</sup>

표본외 예측의 경우에도 비관측요인 모형의 예측력은 상대적으로 우수한 것으로 나타나고 있다. RMSFE 기준으로 1,2,3개월의 예측시계에서, 그리고 MAFE 기준으로는 2개월의 예측시계에서 ARIMA 모형이나 IGARCH 모형에 비해 예측력이 다소 낮게 나타나고 있으나, 예측력에 차이가 없다는 귀무가설은 10%의 유의수준에서 기각되지 않는다. 나머지 모든 예측 시계에서는 비관측

<표 10> 예측력 비교: ARIMA, IGARCH, RS, UC

예측시계		1개월	2개월	3개월	6개월	12개월	24개월	36개월
표 본 내		(RMSFE)						
	UC-0	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	ARIMA(2,1,2)	1.097***	1.094***	1.093***	1.116**	1.183*	1.387*	1.518
	ARIMA(2,1,1)-IG	1.109***	1.109***	1.107***	1.118**	1.177**	1.386*	1.535
	ARIMA(1,1,0)-RS	0.727**	0.760**	0.762**	0.850	0.909	1.050	1.170
		(MAFE)						
	UC-0	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	ARIMA(2,1,2)	1.171***	1.114***	1.110***	1.081*	1.122*	1.352**	1.469**
ARIMA(2,1,1)-IG	1.188***	1.130***	1.132***	1.082*	1.138**	1.351**	1.520**	
ARIMA(1,1,0)-RS	0.892*	0.876**	0.850**	0.847*	0.910*	1.081	1.112	
표 본 외		(RMSFE)						
	UC-0	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	ARIMA(2,1,2)	0.972	0.981	0.997	1,028	1,051	1,105**	1,131**
	ARIMA(2,1,1)-IG	0.977	0.985	0.997	1,025	1,049	1,103**	1,135***
	ARIMA(1,1,0)-RS	1,026	1,056***	1,084**	1,154**	1,190*	1,246*	1,286*
		(MAFE)						
	UC-0	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	ARIMA(2,1,2)	1,031	0,996	1,017	1,019	1,031	1,125*	1,148**
ARIMA(2,1,1)-IG	1,057**	0,998	1,023	1,017	1,029	1,130*	1,169***	
ARIMA(1,1,0)-RS	1,077***	1,072***	1,085***	1,122**	1,142**	1,246**	1,287**	

주: 각 예측시계별로 예측오차가 가장 작은 경우 진하게 표시하였다. IG는 IGARCH를 의미한다.

25) 예측시계가 길어질수록 상대적 예측력이 높아지는데도 불구하고, <표 10>에서 보는 바와 같이 예측력 차이에 대한 검증 결과는 예측시계가 길어질수록 통계적 유의성이 약화되는 것으로 나타나고 있다. 이러한 결과는 예측시계가 길어지면서 예측치의 불확실성(예측오차의 분산)이 높아지기 때문이다.

요인 모형의 예측력이 높으며, 24개월 이후의 장기에서는 통계적 유의성도 검정되고 있다. 국면 전환모형의 예측력은 모든 다른 모형에 비해 낮은 것으로 나타나고 있다.

## V. 결론

본 연구에서는 ARIMA, IGARCH, 국면전환(RS)모형, 비관측요인(UC) 모형들의 주택가격 예측력을 비교하였다. 1개월부터 36개월까지의 예측시계별로 표본내(in-sample) 예측력과 표본외(out-of-sample) 예측력을 살펴보았으며, 예측력 비교는 RMSFE와 MAFE의 두 가지 기준을 적용하였다. Diebold and Mariano(1995) 검정을 통해 모형들의 예측력 차이가 통계적으로 유의적인가의 여부도 검정하였다.

표본내 예측의 경우, 12개월까지의 단기에서는 국면전환모형이 가장 높은 예측력을 보였으며, 24개월과 36개월의 장기에서는 비관측요인 모형이 가장 예측력이 높았다. 또한 이 두 모형은 ARIMA 모형이나 IGARCH 모형에 비해 모든 예측시계에서 예측력이 높았으며, 대부분 통계적으로 유의적인 결과임이 검정되었다. ARIMA 모형과 IGARCH 모형에서는 ARIMA 모형의 예측력이 다소 높았다.

표본외 예측의 경우에는 비관측요인 모형이 대부분의 예측시계에서 가장 높은 예측력을 보였으며, 국면전환 모형의 예측력은 다른 모형들에 비해 크게 낮았다. ARIMA 모형과 IGARCH 모형은 예측시계별로 상대적 예측력 우월성이 다르게 나타났으나, 그 차이는 크지 않아서 예측력은 비슷하다고 보아야 할 것이다.

본 연구에서 검토한 여러 모형들은 변이(change of specification)가 가능한 모형들이다. 다음과 같은 몇 가지 모형의 변이를 생각해볼 수 있다. 우선 GARCH 모형의 경우, GARCH-M 모형을 검토할 필요가 있다. 예측의 측면에서, 일반적인 GARCH 모형과 ARIMA 모형의 차이는 계수값의 차이에만 기인하게 된다. 추정되는 계수값이 비슷한 경우 예측력의 차이는 작을 수 밖에 없다. GARCH-M 모형처럼 GARCH 효과가 예측방정식에 반영된다면 예측력에 큰 변화를 초래할 수도 있을 것이다. 다만 앞에서 언급했던 것처럼 조건부 분산 과정이 불안정적(unstationary)이기 때문에 예측시계가 길어지면서 조건부 분산이 무한대로 커지는 문제가 발생할 수 있다. 따라서 이러한 효과를 제어할 수 있는 방안이 강구되어야 할 것이다. 국면전환 모형의 경우에는 국면전환이 평균관련 상수항에서만 아니라 다른 계수에도 함께 나타나는 일반화 모형이 고려될 필요가 있다. 또한 국면전환 모형의 표본외 예측력이 크게 낮아지는 이유가 미래 국면에 대한 오분류(misclassification of future regimes)에 기인하는 점을 감안한다면, 미래 국면에 대한 예측 방식의 개발도 필요할 것으로 보인다. 비관측요인 모형에서는 추세요인이 임의보행 과정으로 정의되면서 예측은 순환 요인의 시간적 상관성만을 통해 이루어지고 있다. 추세요인이 임의보행 과정이 아니라 예측이 가능한 과정을 포함하게 되면, 비관측요인 모형의 예측력이 높아질 수 있을 것으로 기대된다.

논문접수일 : 2014년 11월 13일

논문심사일 : 2014년 11월 19일

게재확정일 : 2014년 12월 30일

## 참고문헌

1. 김근용, “주택가격 예측을 위한 모형설정과 검증,” 「국토」, 통권 197호, 국토연구원, 1998, pp.54-61
2. 박천규 · 이영, “주택시장 체감지표의 주택시장 지표 예측력 분석,” 「부동산학연구」 제16집 제1호, 한국부동산분석학회, 2010, pp.131-146
3. 손경환 · 지대식 · 최수 · 김승중 · 박천규 외, 「부동산시장의 선진화를 위한 기반정비 연구」, 국토연구원, 2004
4. 손정식 · 김관영 · 김용순, “부동산가격 예측모형에 관한 연구,” 「주택연구」 제 11권 제 1호, 한국주택학회, 2002, pp.49-75
5. 윤주현 · 김혜승, 「주택시장 경기동향 및 단기전망 연구」, 국토연구원, 2000
6. 이영수 · 박상현, “우리나라 주택가격의 추세-순환 분해 및 평균회귀 현상 분석,” 「부동산학연구」 제19집 제4호, 한국부동산분석학회, 2013, pp. 41-54
7. 임재만, “주택매매가격의 변동성에 관한 연구,” 「주택연구」 제 14권 제 2호, 한국주택학회, 2006, pp.65-84
8. 한용석 · 이주형 · 한용호, “지역별 주택가격의 변동성에 관한 연구,” 「대한부동산학회지」 제 28권 제 2호, 대한부동산학회, 2010, pp.9-27.
9. 황영진, “동태적 모형 선택법을 이용한 주택가격 예측변수 분석,” 「부동산학연구」 제19집 제4호, 한국부동산분석학회, 2013, pp.5-26
10. Bessec, M. and O. Bouabdallah, “What causes the forecasting failure of Markov-switching models? A Monte Carlo study,” *Studies in Nonlinear Dynamics and Econometrics* 9-2, 2005
11. Bollerslev, T., “Generalized autoregressive conditional heteroskedasticity,” *Journal of Econometrics*, 52, 1986, pp.307-327
12. Brown, J.P., K. Song, and A. McGillivray, “Forecasting UK House Prices: A Time Varying Coefficient Approach,” *Economic Modeling* 14, 1997, pp.529-548
13. Case, K, and R. Shiller, “The efficiency of the market for single-family homes,” *American Economic Review*, 79, 1989, pp.125-137
14. Clark, P.K., “The Cyclical Component of U.S. Economic Activity,” *The Quarterly Journal of Economics*, 102, 1987, pp.797-814
15. Clark, S. and T. D. Coggin, “Trends, Cycles and Convergence in U.S. Regional House Prices,” *Journal of Real Estate Finance Econ*, Vol. 39, 2009, pp.264-283
16. Crawford, W. and M. Fratantoni, “Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices,” *Real Estate Economics* 31-2, 2003, pp.223-243
17. Crawford, W. and E. Rosenblatt, “Efficient mortgage default option exercise: Evidence from loan loss severity,” *The Journal of Real Estate Research* 10, 1995, pp.543-555
18. Diebold, F. X. and R. S. Mariano, “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, vol 13, no.3, 1995, pp. 253-63.
19. Engle, R.F., “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation,” *Econometrica*,

- 50, 1982, pp. 987-1007
20. Guirguis, H., C. Giannikos, R. Anderson, "The US Housing Market: Asset Pricing Forecasts Using Time Varying Coefficients," *Journal of Real Estate Finance and Economics*, 30:1, 2005, pp.33-53
21. Gupta, R., A. Kabundi, and S. Miller, "Forecasting the US Real House Price Index: Structural and Non-Structural Models with and without Fundamentals," *Economic Modelling*, 28-4, 2011, pp.2013-2021
22. Hall, S., Z. Psaradakis, and M. Sola, "Switching Error-Correction Models of House Prices in the United Kingdom," *Economic Modelling* 14, 1997, pp.517-527
23. Harvey, A. C., "Trends and Cycles in Macroeconomic Time Series," *Journal of Business and Economic Statistics* 3, 1985, pp.216-227
24. Harvey, D., S. Leybourne, P. Newbold, "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, vol 13, 1997, pp. 289-291
25. Hamilton, J.D., "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica* 57, 1989, pp.357-384
26. Malpezzi, S., "A Simple error Correction Model of House Prices," *Journal of Housing Economics*, Vol. 8, 1999, pp. 27-62
27. Mills, W., "Boom-Bust Cycles and the Forecasting Performance of Linear and Non-Linear Models of House prices," *Journal of Real Estate Financial Economics* 36, 2008, pp. 249-264
28. Morley, J.C., C.R. Nelson, E. Zivot, "Why are Beveridge-Nelson and unobserved- component decompositions of GDP so different?" *Review of Economics and Statistics* 85, 2003, pp. 235-243
29. Muellbauer, J. and A. Murphy, "Booms and Busts in the U.K. Housing Market," *Economic Journal* 107(445), 1997, pp.1701-1727
30. Spiegel, M, "Housing Return and Construction Cycles," *Real Estate Economics* 29(4), 2001, pp.521-551