

Data Scientist Handbook

Dookyung Kim

2018-03-15

Contents

Abstract	5
1 Introduction	7
1.1 1안	7
1.2 2안	7
2 이론 및 선행연구 고찰	9
2.1 이론 고찰	9
2.2 선행연구 고찰	10
3 벤치마크 아파트가격지수 추정방법	13
3.1 개 괄	13
3.2 [1단계] 동일주택군 선정 및 실거래 중위가격 추정	13
3.3 [2단계] 동일주택군별 평균 주택가격 추정	15
3.4 [3단계] 서울지역 아파트가격지수 추정(합산)	15
3.5 기존 주택가격지수 추정방법과의 차이	16
4 실증분석 결과	17
4.1 동일주택군 단순 중위가격	17
4.2 동일주택군별 평균가격 추정	17
4.3 합산지수 추정 결과	17
4.4 기존 아파트가격지수와의 비교	17
5 결론	19
A 칼만필터링 A	21
국문요약	23

Abstract

In this paper, we estimated and individual apartment's sales price and its volatility employing a state space model estimated with Kalman filtering and the generalized autoregressive conditional heteroskedasticity(GARCH) model. Our estimation strategy effectively controls for idiosyncratic characteristics in each apartment and also alleviates underestimation problem prevalent when aggregate house price indices are used to estimate price volatility. Empirical results show that the estimated apartment prices fairly follow the actual transaction prices regardless of the volume of transactions. During the period of 2011~2016, house price volatility in Seoul was estimated to be 9.6% on average, which is 8%p higher than the volatility estimated using the data from MOLIT and KB house price indices. For individual apartments, price volatility widely varies ranging from 0% to 28.8% depending on location, size, and construction year. The older and the smaller the individual apartment, its price volatility is estimated to be higher. Apartments located in Gangnam Gu, Nowon Gu, and Yangcheon Gu reveal relatively higher price volatility than those in the other regions in Seoul. The relationship between price volatility and transaction volumes or the rate of change in house price have weakened since 2010, and the latter two variables tend to lead the volatility by two to three months.

Chapter 1

Introduction

1.1 1안

전세계 주택가격은 2008년 글로벌 금융위기 직후 큰 폭 하락하였다가 2011~12년을 저점으로 다시 급반등하여 최근에는 대부분의 국가에서 사상 최고치를 다시 상회하였다. 특히 전세계적으로 부동산과 관련된 금융부채 다시 급속히 증가하고 있어 향후 부동산 경기에 대한 기대와 전망이 경제와 금융시장의 미치는 영향이 날로 커지고 있으며, 중앙은행들도 통화신용정책도 주택시장의 가격변동성에 크게 제한을 받고 있다. 최근 주택가격지수 추정에 관한 연구는 크게 2가지 흐름을 가지고 있다. 첫째, 전세계적으로 비교가능한 지수를 개발하려는 노력이다. IMF, BIS 등에서는 부동산 자산의 변동성이 금융시스템의 안정성을 저해하는 주요 요인으로 인식하고 2010년 이후 전세계 국가간 비교가능하고 주택시장의 상황을 정확히 반영하는 주택가격지수의 개발을 위한 연구가 진행중이다. 주택가격지수와 관련한 전세계적인 현상 중 하나는 다양한 지수의 개발 필요성에 따라 목적별로 다양한 지수가 개발되고 있다. 반복매매지수, 헤도닉지수, 평가기반지수, 가중평균지수 등 ... 부동산은 개별성과 거래부족 등의 문제로 가격지수 추정이 근본적으로 한계가 있는 데다 각국의 데이터 이용가능성이 매우 상이하여 비교가능한 주택가격지수 추정에는 한계가 있기 때문에 목적에 맞는 다양한 지수를 개발하고자 하는 이유. Borio와 Lowe(2002)는 부동산 가격지수의 다양성을 중요하다고

우리나라에서는 국민은행과 일부 부동산업체에서 1995년 이후 평가가격에 기초한 주택가격지수를 발표하고 있으며 2006년부터는 국토교통부에서는 실거래가가격에 기초한 주택가격지수를 발표하고 있다. 그러나 우리나라의 경우 주택시장의 실거래가가격이 15년 이상 발표되고 있는 데다 주택시장이 아파트 시장 중심이어서 주택의 동질성이 높고 전산화와 IT발전으로 주택시장과 관련하여 이용가능한 자료도 비교적 많고 국민들의 부동산에 대한 관심도 높음에도 불구하고 주택과 부동산시장에 대한 다양한 가격지수의 개발은 거의 이루어지지 않고 있는 점은 다소 의문이다. 기존 선행연구는 데이터 확보가 용이한 반복매매지수모형을 중심으로 한 연구가 대부분이며 주택특성을 기반으로 하는 헤도닉모형 기반의 가격지수 연구는 거의 이루어지지 못하고 있다. 국토교통부의 실거래가지수도 지수작성의 권역이 넓어서 부동산금융의 실무담당자가 이용하기에는 다소 한계가 있다.

1.2 2안

주택가격지수는 ...로써 ... 등에 광범위하게 활용되는 주요 경제지표 중의 하나이다.

주택은 가계, 국가차원에서 가장 중요한 자산중이 하나. 그러나 주택이 가격지수를 추정하는 일은 매우 어려운 일. 이는 부동산의 경우 개별성이 강하고 거래가 많지 않은 데다 주택가격지수를 추정하는데 필요한 정보도 충분하지 않은 경우가 대부분이기 때문이다. 이에 따라 주택가격지수를 추정하려는 노력이 국내외적으로 활발히 진행중

우리나라에서는 국민은행 ...와 2006년부터는 국토교통부 실거래가 지수와 실거래 내역자료를 공개하고 있다. 그러나 현행 주택가격지수는 지수작성의 권역이 넓고 ... 해서 부동산 금융의 실무담당자가 이용하기에는 한계가 있다(박연우 & 방두완,). 현행 주택가격지수는 행정구역 단위로 편제되어 해당 지역내에 평균적인 주택의 가격동향에 대한 정보만을 줄 뿐이며 주택시장에 영향력이 크고 전체 시장에 대해 선행성이 있는 대형 아파트 단지의 가격정보를 제공하지 못한다. 반면 주식시장에서는 상장기업의 주가를 종합한 KOSPI지수가 시장을 대표하는 벤치마크지수로서 주식 실물거래 뿐아니라 선물, 옵션 등 다양한 파생상품의 벤치마크지수로 활용되면서 관련 시장 활성화에 많은 기여를 하고 있다. 코스피 시장에 상장된 기업은 2017년말 현재 1천 개로 전체 기업 1백만개의 %에 불과하나 국민경제에서는 차지하는 비중은 %에 달한다는 점에서 상장기업의 주가를 대표하는

종합주가지수가 한국을 대표하는 종합주가지수로 자리매김하고 있다. 벤치마크는 성과평가지표 시장5% 상승시 보유주식이 %상승 비교

주택시장에서는 이러한 벤치마크지수가 존재하지 않음 단지특정 아파트단지나 거래량이 많은 대형단지¹를 전체 시장의 평균

그러나 현행 주택가격지수는 포괄범위가 넓어 실무에 활용하는데 제한. 국제적으로도 목적에 맞는 다양한 주택가격지수를 개발하고자 하는 노력이 진행중. 국내에서도 실거래가 지수가 발표된 지 10여년이 지난 시점이므로 다양한 주택가격지표를 산출할 필요

이에 본 연구에서는 주식시장에서 상장기업의 주가를 대표에서의 종합주가지수(KOSPI)와 같은 서울 아파트시장에서 거래량이 많은 아파트, 특히 대형 아파트의 거래상황을 대표 주택시장의 벤치마크 가격지수를 개발하고자 한다. 여기서는 기존 선행연구에서 제시한 주택가격지수가 가져야할 특징과 더불어 금융시장에서 벤치마크지수가 가져할 속성을 모두 가지고 있는 주택가격지수를 도출하는 것으로 목적으로 한다. 특히 동 지수에 포함된 아파트들의 개별 가격지수도 각기 제공 국제적으로도 다양한 주택가격지수의 개발 필요성을 제기

시장에서 신뢰할 만한 주택가격지수를 만들 수 있다면 주택 투자시 참고지수, 보유주택의 종합수익률 측정, 금융, 정책입안 등 부동산과 관련한 다양한 실무에서 활용이 가능할 것으로 기대된다.

본 연구에서는 실무적으로 사용이 가능하고 주택 관련 투자의 벤치마크로 활용 가능한 주택가격지수를 개발하고자 한다. 벤치마크란 ...

본 연구는 실무에서 적용가능하고 기존의 주택가격지수 추정방법의 한계를 극복하는 보다 나은 주택가격지수 추정방법을 제시하고자 한다. 전세계적으로 통계당국에서 실무적으로 활용하는 지수는 반복매매지수, 헤도닉 방법론은 이론적으로는 우수하지만 정보부족으로 실무에 활용하지 못하고 있다. 그러나 우리나라의 경우 아파트시장 중심의 동질성이 높은 주택시장의 특성과 아파트와 관련한 정보가 인터넷 업체 등에서 제공하는 정보도 풍부한 편이므로 이를 활용하여 보다 나은 주택가격지수를 개발하는 것이 여타 국가에 비해 상황이 좋다.

본 연구에는 실무에 적용가능한 헤도닉 기법과 시가총액 기반의 주택가격지수 추정방법을 개발하기 위해 일차적으로 상태공간모형과 칼만필터링 기법을 활용하여 강남지역 몇 개 개별 아파트 단지의 평균 실거래 가격을 매월 추정하였다. 아울러 이를 기반으로 동 아파트를 대상으로 헤도닉 기반의 주택가격지수를 추정하였다. 개별 아파트 가격의 집계는 아파트의 시장가치를 반영하기 위해 Weighted 방식으로 추정 아울러 Arthur(BIS, 2003)가 제시한 이상적인 부동산가격지수의 6가지 특성(regular availability, representativeness, comparability, continuity, length of series, frequency, parameter stability, index stability)의 측면에서 본 연구가 제시한 추정방법과 기존의 주택가격지수 추정방법을 비교하고자 한다.

이러한 배경에서 본 연구는 국토교통부가 공개하는 아파트 실거래 자료(2006~2016년도)를 이용하여 서울지역 개별 아파트의 단지/평형별 가격을 추정하고 이를 기초로 개별 아파트 가격(단지/평형별)의 변동성을 추정하는 새로운 방법론을 제시하고자 한다. 이를 위해 금융경제분야에서 널리 활용되고 있는 상태공간모형과 GARCH모형을 개별 아파트 가격과 그 변동성 추정방법으로 활용하였다.² 개별 아파트 가격 추정에는 Hinrichsen and Holmes(2009)가 발표한 R package인 KalmanEM(Kalman Filter & EM algorithm)을 이용하였다. 상태공간모형과 GARCH모형은 부동산 분야에서도 많이 활용되고 있으나 이를 조합하여 개별 아파트 가격의 변동성을 추정한 사례는 아직까지 없었던 것으로 생각된다.

본 연구는 다음과 같이 구성된다. 제Ⅱ장에서는 자산가격의 변동성에 관한 이론과 선행연구를 고찰한다. 제Ⅲ장에서는 개별 주택가격과 변동성 추정방법을 설명하고, 제Ⅳ장에서는 추정결과를 제시한다. 마지막 제Ⅴ장에서 본 연구의 의의와 한계, 그리고 향후 연구방향을 제시한다.

¹ star 지수

² Schulz and Werwatz(2004)와 박현수·유은영(2014)은 상태공간모형을 활용하여 거래빈도가 작은 하부지역의 주택가격지수를 추정한 바 있다.

Chapter 2

이론 및 선행연구 고찰

2.1 이론 고찰

경제학원론에서 재화의 가격과 거래량은 수요와 공급에 의해 주로 결정되는 내생변수로 본다. 즉 모든 재화는 이를 둘러싼 경제전반의 외생요인 변화에 따라 수요와 공급 여건이 변하게 되며 이로 인한 시장의 불균형 발생과 균형으로의 회귀가 반복되는 과정에서 가격과 거래량이 상호 연관되어 변하게 된다. 경제주체들은 합리적인 경제활동을 영위하기 위해 이러한 가격과 거래량의 움직임을 정확히 측정하고 예측하기 위해 노력하고 있다. 시장에서 거래되는 모든 재화는 이론적으로 가격이 존재하지만 개별 재화의 가격을 가격지수¹로 산출하는 것은 쉬운 일이 아니다. 특정 재화에 대한 합리적이고 정확한 가격지수를 산출하기 위해서는 우선 해당 재화에 대한 잠재적인 수요와 공급이 풍부하게 존재하고 많은 거래량이 꾸준히 발생하여야 한다. 또한 모든 거래정보가 집중되고 공개되어야 한다. 그러나 이러한 경우는 경제학에서 이론적으로 가정하는 완전경쟁시장²에서나 가능하며 현실 경제에는 이러한 재화는 매우 이례적이 경우에만 존재한다. 이러한 이유로 모든 통계기관들은 가격지수를 산출하는 과정에서 다양한 추론방법과 가정을 활용하고 있다. 대표적인 가격지수의 예로는 소비자물가지수, 국제원유가격, 주가지수, 부동산가격지수 등이 있다. 각 가격지수 유형별로 지수산정시 어려움과 해결방안을 살펴보면 다음과 같다. 첫째, 소비자물가지수와 같이 다양한 상품군의 가격지수를 산출하는 경우 대상 상품군의 선택, 각 상품이 가격지수에서 차지하는 비중(sample selection), 각 상품가격으로부터 전체 상품군의 가격으로의 합산방식(aggregation), 각 상품별 가격측정방식 등과 관련하여 추가적인 정보와 의사결정이 필요하다. 물가지수의 경우 각국의 통계당국이 각국의 상황에 맞추어 결정한 편제방식으로 추정하고 있으며, 국제적으로 표준화하기 위해 ..에서 매뉴얼을 만들어 제공하고 있다. 둘째, 국제원유가격과 같이 대상자산이 한 가지인 경우에도 재화의 동일성 유지와 거래정보를 입수하는데 어려움이 있을 수 있다. 원유는 화학적 구성의 차이로 그 품질이 시기별, 지역별로 모두 다르며 전세계에 걸쳐 다양한 가공단계에서 거래되는 원유의 가격과 거래량을 모두 취합하여 가격지수를 산출하는 것은 사실상 불가능하기 때문이다. 이러한 문제를 해결하기 위해 원유의 성분 등을 고려하여 품질등급을 사전에 정해두거나 지역별로 시장을 분할하여 가격지수를 산출할 수 있다. 국제원유시장에서는 두바이와 WTI 유가가 이러한 이유로 서로 다른 가격지수로 산출되어 거래되고 있다. 셋째, 엄격하게 조직화된 금융시장에서 거래되는 주식

둘째, 주가지수와 같이 금융자산의 가격 - 거래정보가 완전 공개, 품질이 동일 -> 경제학에서 가정한 완전경쟁시장과 가장 유사 - 자산가격변화에 대한 기대 - 공시제도 등 엄격한 제한. - 상장된 주식만(대상 제한) 전체주식이 불과 %

넷째, 부동산시장 그러나 주식시장과 같이 조직화된 시장이 존재하지 않는 경우 - 어려운 이유

주택가격 산정이 어려운 이유 : 타 상품에 비해 개별성이 강한 반면 거래량은 매우 부족 -> 동일한 주택이 거래되는 사례가 거의 없음, 거래정보와 주택정보를 매번 아는 것이 어려움

- 통계적 추론방법에 주로 의존.
- 대상이 모호 -> 실무적 활용이 어려움
- 목적, 샘플편의 방대한 정보 필요
- 그러나 개인 부에서 차지하는 비중이 가장 큰 자산
- 보다 실무적으로 활용가능한 지수 필요(소비자물가지수 연금 등, 주가지수 연계상품...) 최근 국제기구 등에서 연구가 활발

주요 요건 정의 ... 다양성, 실무에 적용가능한 지수

¹ 가격지수란 특정 재화의 가격을 특정 시점을 기준으로 지수화한 수치를 의미한다. 만약 사과 1개의 가격이 기준년도에 1000원이었는데 현재 1400원이라면 가격지수는 기준년도의 가격을 100으로 보아 140이 된다.

² 다수의 수요자와 공급자, 재화의 동질성, 자유로운 진입과 퇴거, 완전한 정보(가격, 거래량, 재화정보 포함)

시장에서 거래되는 모든 상품은 통상적으로 가격이 존재하며 특정 상품군의 가격지수가격이 존재하며 세상에 존재하는 모든 재화는 이론적으로 특정 상품군의 가격지수를 편제하는 일은 생각보다

개별상품의 가격은 각기 특수한 수요·공급의 사정에 따라 그 변동의 크기와 방향이 일정하지 않다. 두 시점 사이에서 한 상품의 가격은 등귀(騰貴)하나 다른 상품의 가격은 하락하는 경우도 있으며, 같은 방향으로 등락하는 경우라도 반드시 같은 비율로 변화하는 것은 아니다. 따라서 국민경제 전체의 물가변동을 관찰하기 위해서는 가격지수가 아니라 물가지수가 이용된다.

주택은 가계, 국가차원에서 가장 중요한 자산중이 하나. 그러나 주택이 가격지수를 추정하는 일은 매우 어려운 일. 이는 부동산의 경우 개별성이 강하고 거래가 많지 않은 데다 주택가격지수를 추정하는데 필요한 정보도 충분하지 않은 경우가 대부분이기 때문이다. 이에 따라 주택가격지수를 추정하려는 노력이 국내외적으로 활발히 진행중

주택가격지수 추정에 관한 연구는 크게 3가지 측면에서 활발히 이루어지고 있다.

첫째, 정확한 주택시장의 흐름을 이해하기 위해

가 보유하는 자산중 그 비중이 가장 큰 자산, 국가차원에서는 적으로 보더라도 매우 중요한 자산이다. 이며 국가계가 보유하고 있는 자산 중 80% 이상이 부동산이며 이중 절대적개인과 국가가 가지는 전체 총자산 중 주택이 차지하는 비중은 절대적으로 크다. (부)부중 의 부에서 차지하는 비중이 절대적으로 크주택가격지수는

2.2 선행연구 고찰

순수한 가격변화 추정을 위한 고려사항

- Quality Mix change
 - 개별주택의 속성은 변함.(주택개량, replacement, 증축, depreciation)
 - Compositional change : 거래된 주택의 구성이 시기별로 변함(경기침체에 대형주택 거래 감소)
- major characteristics : 주택면적, 토지면적, location, age, type of structure, material used, 방수, 화장실수, 창고, 수영장, 에어컨, 편의시설
- 주택 특징 변화를 control하는 방법 : stratification or mix adjustment, repeat sales mothds, hedonic regression method, use of property assessment information

2.2.1 헤도닉 지수(Hedonic Price Index)

헤도닉가격지수는 시장에서 거래된 주택의 거래가격을 종속변수로 하고 주택가격에 영향을 미치는 여러 특성³과 기간더미를 설명변수로 설정한 회귀모형을 통해 주택의 특성변화 따른 가격변화를 제외한 ‘질적으로 동일한 주택’의 순수한 가격변화를 추정하는 방법이다(Rosen, 1974).

$$\ln P = \alpha + \sum_{i=1}^k \beta \ln X_i + \sum_{i=2}^T \gamma_i D_i + \epsilon \quad (2.1)$$

$P :$

$X_i :$

$D_i :$

헤도닉 회귀분석은 주택의 질적 변화, 표본선택 편의를 고려할 수 있다는 장점이 있으나 주택의 특성요인 정보 등 많은 데이터를 필요(data-intensive)로 하며 주택의 위치를 통제할 수 없다는 단점이 있다. 동 방법은 미국상무부 통계청의 신규단독주택가격지수와

³국내 선행연구에서는 대지면적, 층수, 준공연수 등의 구조특성 변수, 지하철역과의 거리, 도심으로부터의 거리, 접면도로의 폭 등의 입지특성 변수, (임대료를 추정하는 모형에서의) 임대기간, 입주자의 신용도 등의 계약 특성변수 등을 사용하고 있다.(이용만·이상한 (2008), 김명준 외 2008, 박현수 2009) (location, type, age, floor space, plot area만으로 충분히 설명)

영국통계청의 주택가격지수, 프랑스에서는 전지역의 주택특성가격지수 작성시 활용되고 있으며 주택의 질적 변화를 통제할 수 있는 가장 이상적인 가격지수 추정방식으로 평가되어 학계에서 많이 활용되고 있다.

2.2.2 가중평균가격지수(Stratification or Mix Adjustment Methods)

가중평균가격지수는 주택을 위치, 규모, 종류, 노후도 등 유형별로 세분화하여 각 그룹의 단위면적당 평균가격 또는 중앙값⁴을 산출한 후 이를 다시 거래량 또는 주택수(stock)를 기준으로 가중평균하여 주택가격지수를 추정하는 방법이다(CPI Manual, 2004)⁵. 동 방식은 주택의 유형별, 지역별 가격지수를 쉽게 산출할 수 있는 데다 주택의 그룹을 잘 정할 경우 매 기간 거래된 주택의 구성이 달라져도 각 그룹의 관측치 수만 달라지고 평균가격은 크게 영향받지 않는 장점이 있다. 다만 주택의 그룹구분이 적정하지 못하거나 너무 세분된 경우에는 표본선택 또는 단위가격 편이, 표본부족에 따른 높은 가격변동성 문제 등이 발생할 수 있는 단점이 있다. CPI Manual(2004)은 동 방법 활용시 각 계층의 평균값은 단위면적가격(unit value)의 중위값을 사용하고, 합산은 연쇄가중 피셔가격지수(chained Fisher index)를 사용할 것을 권고하고 있다.

Laspeyres price index

동 방법은 영국 환경부(1982)와 호주 통계청(ABS, 2006) 등 일부기관에서 주택가격지수 편제방법으로 활용하고 있으나 학계에서는 여타 방법론과의 비교목적으로 주로 활용되고 있다⁶.

2.2.3 반복매매지수(Repeat Sales Price Index)

반복매매지수는 전체 지수산정 기간중 2번 이상 거래된 동일주택의 반복매매거래상 자료(주소, 매매가격, 거래일 등)로부터 해당기간 가격상승율을 기간 더미변수에 대해 회귀분석하여 주택가격지수를 산출하는 방법이다(Bailey, Muth & Nourse, 1963). 동 방법은 동일주택의 반복매매거래상 정보만을 이용하므로 주택의 위치정보를 자동으로 통제하며 데이터의 수집부담이 적어(less data-intensive) 실무적으로 이용하기 용이하다. 다만 주택거래량이 많지 않은 경우 표본선택편의가 발생하고 주택의 질적 변화를 고려하기 어렵다. 또한 전체 주택거래자료 중 두 번 이상 거래된 주택거래자료만 이용하므로 데이터 이용의 효율성이 낮다는 단점이 있다. ...에서 활용중. 특히 Case & Shiller(1987, 1989)는 각 반복매매거래상의 거래일 간격이 길어질수록 주택가격의 변동성이 커지는 이분산(Heteroscedasticity) 문제를 해결할 수 있는 가중회귀분석(Weighted Least Square) 방법론을 제시하였다. 동 Case-shiller지수는 미국의 대표적인 주택가격지수로 발전하여 최근에는 이에 기초한 선물상품이 도입되어 거래되고 있다.

```
knitr::include_graphics("images/figure25.png")
```

2.2.4 평가가격기반 지수

2.2.5 상태공간모형 기반

최근의 주택가격지수 추정에는 주택거래에서 실제로 관측되는 거래가격에 내재한 오차(transaction noise)를 제외한 주택가격지수를 산출하려는 목적에서 주택의 시장가격을 잠재변수(latent variable)로 설정하는 상태공간모형을 접목한 연구들이 로 설정하여 가격지수를 추정하는 상통제하여 주택의 시장가격을 추정하려는

2. latent variable : selling prices do not always exactly represent the market values of the properties, which can be viewed as a latent variable. -> transaction noise involved that causes volatility of the measured price indices.

주택가격지수 추정에 상태공간모형을 접목한 연구들은 흔히 않다. Schulz and Werwatz(2004)는 1982~1999년 기간 동안 독일 베를린 남서부 4개 지역의 실거래가 자료에 기초하여 기존의 헤도닉가격 함수를 이용하여 주택가격 움직임의 공통 요인(common price component)을 찾고 이를 통해 주택가격지수를 산출하는 방법이 가지고 있는 한계를 극복하기 위해 상태공간모형을 이용하였다. 그들은 자산가격결정에 대한 이론적 모형을 통해 특정 주택의 가격과 기준 주택 가격의 차이가 해당 주택이 가지고 있는 특성뿐 아니라 해당 주택의 기대수익률에 의존함을 제시하고 주택가격과 기대수익률 결정요인을 동시에 추정하는 과정에서

⁴CPI Manual(2004)에서는 주택가격의 positively skewed된 경우가 많아 mean보다 중위값이 많이 사용unit price 중위가격을 사용토록 권고

⁵이러한 'stratification and re-weighting' 방법은 표본데이터를 가지고 분석하는 경우 표본선택편의(sample selection bias), 미응답(non-response) 등에 따른 문제를 해결하는 일반적인 방법론이다.

⁶Mar and Goldberg(1984), Crone and Voith(1992), Gatzlaff and Ling(1994), Wang and Zorn(1997) 등

$$1. \text{ Index Coefficients: } \hat{\beta} = (D'D)^{-1}D'r; \quad r_t^i = \log\left(\frac{P_t^i}{P_{t-1}^i}\right) = \sum_{t=1}^T \beta_t D_t^i + \varepsilon_t^i$$

$$2. \text{ Weighted regression: } \varepsilon^2 = c + HP; \quad \varepsilon_t^i = \beta_0 + \sum_{i=1}^I \beta_1 H P_t^i + \mu_t^i$$

$$3. \text{ GLS: } \tilde{\beta} = (D'W^{-1}D)^{-1}D'W^{-1}r; \quad \tilde{r}_t^i = \sum_{t=1}^T \tilde{\beta}_t \tilde{D}_t^i + \tilde{\varepsilon}_t^i$$

Figure 2.1: Case-Shiller Index

시점별로 거래빈도가 서로 다른 문제를 극복하기 위해 상태공간모형을 이용하였다. 국내에서는 박현수·유은영(2014)이 거래빈도가 낮거나 거래가 없는 하위 시장을 대상으로 상태공간모형을 활용하여 실거래가격지수를 개발하고, 이를 반복매매모형에 기초한 실거래가격지수와 비교 분석한 바 있다. 이러한 선행연구들은 시점별로 거래빈도가 서로 다른 경우 발생할 수 있는 문제를 중심으로 상태공간모형을 이용하여 주택가격지수를 개선하는 것을 목적으로 하였기 때문에 관련 자료를 지역별로 통합하여 이용하였다.

Chapter 3

벤치마크 아파트가격지수 추정방법

3.1 개 괄

본 연구에서는 기존 주택가격지수가 가지는 ... 등의 문제를 개선하고 실무적으로 쉽게 적용할 수 있는 실거래가 기반의 개별 아파트 평균매매가격 및 다양한 유형 그룹별 아파트 가격지수를 체계적으로 추정할 수 있는 새로운 방법론을 제시하고자 한다. 특징 : 샘플대상 주택별로 월별 가격지수 산출, 시장성이 있는 주택을 대상 이를 위해 본 연구에서는 ...만을 분석대상으로 하였다.

본 연구에서는 다음과 같이 3단계에 걸쳐 새로운 아파트가격지수를 추정한다. 1단계에서는 국내 선행연구에서 반복매매지수 추정방법 적용시 사용하는 동일주택(동일단지, 동일평형, 동일층 아파트) 개념을 일부 완화하여 동일주택군(동일단지, 동일평형 아파트)이라는 개념을 도입하고 동일주택군별 월별 평균매매가격(중위값)을 추정한다. 이 경우 동일주택군내 반복매매거래쌍 자료가 크게 증가하는 장점이 있으나 분석대상 아파트의 동일성이 약화되어 실거래가격에 오차(transaction noise)가 증가하는 단점이 있다. 이에 2단계에서는 동일주택 개념 확대로 발생하는 측정오차를 통제하는 한편 동일주택군 내에서 주택거래가 한 건도 발생하지 않는 경우 발생하는 누락치를 추정(imputation)하는 방법으로 상태공간모형과 칼만필터링기법을 활용하였다. 이를 통해 전체 동일주택군의 평균매매가격을 매월 누락없이 추정하는 것이 가능하다. 3단계에서는 기존 선행연구에서 활용하는 다양한 합산방법을 활용하여 동일주택군별 주택가격에서 시장성이 있고 벤치마크로 활용가능한 서울지역 아파트가격지수와 다양한 유형그룹별 주택가격지수를 추정하였다.

3.2 [1단계] 동일주택군 선정 및 실거래 중위가격 추정

주택가격지수는 기본적으로 일정기간중 동일한 주택의 가격변화를 측정하는 것을 목적으로 한다. 그러나 주택시장은 개별성이 강하고 거래량이 매우 제한적이어서 현실적으로 동일한 주택의 가격변화를 측정할 수 있는 자료가 매우 제한적이다. 이를 감안하여 대부분의 주택가격지수 추정방법들은 동일주택의 가격변화 관련 기초자료 확보 또는 유형별 주택그룹 구분 등을 위해 동일주택에 대한 가정과 통계적 추론방법을 사용한다. 국내 아파트의 경우 단독주택, 연립주택 등과 달리 주택의 특성이 비교적 표준화되어 있는 데다 동일단지 아파트간 유사성 정도가 높아 동일주택 그룹 구분이 상대적으로 용이하고 명확한 편이다. 기존 선행연구와 국토교통부의 실거래가 지수 산정시에는 “동일 단지, 동일 평형, 동일 층인 아파트”를 동일주택으로 가정하여 지수를 산출하고 있다. 이 경우 아파트들을 작게는 몇 세대에서 많게는 몇 십 세대까지 동일주택 그룹을 묶는 것이 가능하고 동일주택내 반복매매거래쌍도 제한적으로 확보할 수 있어 행정구역단위의 주택가격지수 산출이 가능해진다. 실제로 선행연구의 동일주택 가정과 국토교통부가 제공하는 2006~2017년중 서울지역 아파트 실거래 매매자료를 활용하여 분석해본 결과, 동 기간중 매매건수는 총 928,325건, 거래실적이 있는 아파트는 총 8403단지¹로 나타났으며 동일주택은 총 몇 개로 조사되었다. 이중 월별로 실제 주택거래가 발생한 아파트는 단지기준으로는 평균 %, 동일주택 기준으로는 %에 불과

그러나 개별 동일주택의 가격지수를 추정하고 이를 활용하여 다양한 유형별 주택가격지수를 추정하는 것은 여전히 불가능하다. 매월 거래되는 아파트가 달라지므로 샘플링 문제도 존재. 본 연구에서는 동일주택 가정을 보다 완화하여 “동일 단지, 동일 평형 아파트”를 “동일주택군”이라고 가정하고 국토교통부에서 제공하는 아파트 실거래 상세자료와 다음부동산의 아파트 단지별 정보를 활용하여 동일주택군을 다음과 같이 도출하였다. 동일주택군의 아파트 세대수가 최소 150세대 이상되는 대형 아파트단지를 분석대상으로

¹ 국토교통부가 공개하는 자료의 주소지를 기준으로 구분

<표 1> 서울지역 아파트 및 본 연구의 대상주택 현황(2016년)

	단지수	동수	세대수	전용면적		
				60 m^2 이하	60~102 m^2	102 m^2 초과
전 체(A)	4,256	19,959	1,518,611	363,610	737,969	417,032
분석 대상(B)	1,347	-	714,998	286,635	346,743	81,620
(B/A, %)	(31.6)	(-)	(47.1)	(78.8)	(47.0)	(19.6)

자료 : 서울 열린데이터 광장, 서울시 아파트 현황 통계

Figure 3.1: 서울지역 아파트 및 본 연구의 대상주택 현황(2016년)

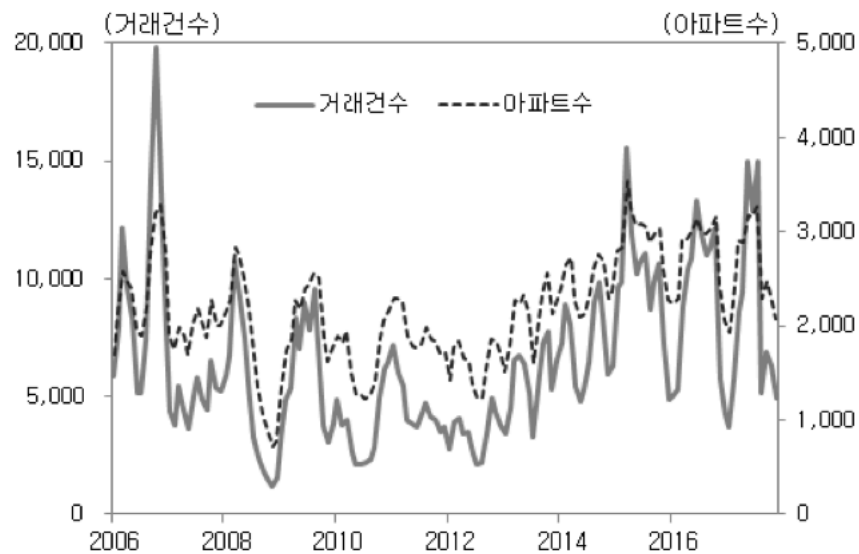


Figure 3.2: 월평균거래건수 및 아파트수

제한하였다. 이를 통해 동일주택군의 주택실거래가 적어도 분기에 한 번 이상은 나올 수 있도록 하였다. 동일주택군은 기존 선행연구에서 분류한 세부기준으로는 몇 개, 본 연구의 기준으로는 몇 개로 나왔다.

은 총 몇 개가 나온다. 그러나 이 경우에도 동일주택군에서 매월 주택거래가 발생하기는 어렵기 때문에 거래가 전혀 발생하지 않는 월이 많이 발생하게 되므로 매월 가격지수를 산출하는 것은 여전히 어렵다.

그러나 세대수가 많은 대형 아파트 단지의 경우에는 전체 샘플기간(144개월)중에 일부 누락치를 외에는 대부분 매월 거래가 발생하고 있음. 이에 본 연구에서의 대상주택은 동일주택군을 대상주택수가 150세대 이상이고 월평균 거래수가 0.5 이상인 아파트로 가정. 이에 해당하는 동일주택군은 총 3000개. 서울 아파트의 대상 동일주택군은 세대수가 150~1000세대까지 분포. 지역별로 보면.

[그림]전체 주택 분포에서 동일주택군 분포

동일주택에 대한 가정에 따라 동일주택의 수를 비교하면 다음과 같음 동일주택군 기준으로 보면 총구분시와 미구분시 전체 동일주택군 수는 몇 개, 월평균 거래건수는 건으로 큰 차이 총 미구분시 동일주택의 거래분포는 다음과 같다. 동 기간중 월평균 거래건수는 6492건, 실제 거래실적이 있는 아파트 단지는 평균 2186개 정도이며 나머지 6천개 단지는 전혀 거래가 발생하지 않는다. 즉 월별로 보면 전체 아파트중 25%만이 거래가 발생하고 거래건수는 아파트별 월평균 3건 정도에 불과

```
knitr::include_graphics("images/figure30.png")
```

```
knitr::include_graphics("images/figure35.png")
```

대상 동일주택군이 선정되었다는 것을 전제로 실거래가 자료를 기준으로 3천개 동일주택군의 중위가격은 개별 동일주택군별로 월별 중위가격을 사용하여 산출. 선행연구에서 중위가격이 우수하다고 증명.

국내 아파트의 실거래가지수를 연구한 선행연구와 국토교통부의 실거래가지수는 국내 아파트시장의 특성을 감안하여 동일단지, 동일평형, 동일층 아파트를 유사 아파트로 가정하고 해당지역내 유사 아파트의 반복매매거래쌍 자료를 활용하여 행정구역별 아파트가격지수를 산출하고 있다. 본 연구에서는 행정구역보다 세분된 아파트별 월평균 매매가격²을 추정하기 위해 유사 아파트 가정을 동일단지, 동일평형으로 완화하였다. 이 경우 유사 아파트의 반복매매거래쌍 자료가 큰 폭으로 증가하는 장점이 있으나 유사 아파트의 동질성이 약화되어 실거래가격에 오차(transaction noise)가 증가하는 단점이있다. 이에 본 연구에서는 실제 관측되는 유사 아파트의 거래가격과 별도로 동일주택의 거래가격을 잠재변수로 가정하는 상태공간모형을 설정하고 칼만필터링 기법을 적용하여 측정오차를 필터링하여 잠재변수인 동일 주택의 거래가격을 추정하였다.³

본 연구에서는 기존 주택가격지수 추정방법론인 가중평균가격지수, 반복매매지수와 상태공간모형과 칼만필터링 기법을 활용하여 벤치마크 아파트가격지수 추정방법을 제시하고자 한다. 1단계에서는 표본편의를 제거하기 위해 동일주택군을 선택. 우리나라의 아파트시장은 표준화되어 있어 동일주택군을 선택하기가 비교적 용이.

-> 대부분의 통계편제 기관들은 제한된 거래정보로 쉽게 가격지수를 편제할 수 있는 반복매매기법을 많이 활용. 실무적으로는 일정한 속성(지역, 크기, 주택유형 등)이 같은 주택군을 동일주택으로 가정하고 경우 동일주택의 반복매매쌍 정보를 활용하여 주택가격지수를 산출 동일주택으로 분류되는 주택의 수가 늘어나면 자연히 더 많은 반복매매쌍이 존재 -> 정기적으로 가격지수 산출이 가능해짐

본 연구에서는 동일주택 가정을 보다 완화하여 동일 단지, 동일 평형 아파트를 동일주택으로 가정하였다. 이로써 동일주택의 수를 150~1000세대까지 확대 가능. 그러나 이 경우에도 동일주택군에서 매월 주택거래가 발생하기는 어렵기 때문에 거래가 전혀 발생하지 않는 월이 많이 발생하게 되므로 매월 가격지수를 산출하는 것은 여전히 어렵다. 아울러 기존 선행연구와 달리 아파트의 층을 고려하지 않았기 때문에 실제 거래금액에 측정오차가 포함되는 문제가 생긴다. 이를 해결하기 위해 상태공간모형과 칼만필터링기법을 활용하여 ... 기법으로 측정오차 및 누락관측치 문제를 해결가능. 이를 통해 가격지수 산출의 대상주택군 전체에 대해 매월 평균가격을 도출 가능

일반적인 가격지수 산출시 개별 품목의 가격에서 전체 품목의 가격지수화를 위해 다양한 합산방법((중위가격, 거래량 가중평균, 시가총액 가중평균 등)을 사용

3.3 [2단계] 동일주택군별 평균 주택가격 추정

동 중위가격에는 해당 동일아파트군내 개별아파트의 특성차이(층, 인테리어) 등에 따라 측정오차를 내재 이를 통제하기 위해 본 연구에서는 .. 사용 상태공간모형과 칼만필터링은 ...

여기서는 위에서 구한 동일주택군별 중위가격을 측정오차를 포함하는 측정변수로 보고 실제 평균가격은 보이지 않는 잠재변수로 가정하여 상태공간모형을 설정하고 칼만필터링기법을 활용하여 잠재변수인 동일주택군의 평균가격을 추정 이를 통해 전체 동일주택군에 대해 누락치가 없는 평균가격을 추정가능 위에서 구한 중위가격에 비해 편차가 줄어든 평균가격이 산출. 이를 통해 층, 인테리어등 개별 주택의 특성요인을 체계적으로 통제 가능

상태공간모형, 칼만필터링 (그림으로 설명)

3.4 [3단계] 서울지역 아파트가격지수 추정(합산)

다음 단계에서는 앞에서 구한 동일주택군별 평균가격을 합산하여 서울지역 아파트가격지수를 추정. 합산방식은 다양한 방법을 사용가능

(시가총액과 거래량 가중평균) 주가지수 등의 산출방법에서 사용하는 시가총액, 거래량 가중평균방식으로 동일 아파트군의 평균가격을 합산

²해당 월에 두 건 이상의 유사 아파트 거래가 있는 경우 중위값을 사용하여 월별 평균가격을 산출하였다.

³선행연구에서는 유사 아파트내 조망권, 남향여부 등의 차이에 따른 거래가격의 측정오차를 전혀 고려하지 않고 있으나 동 연구에서는 유사 아파트의 이러한 추가적인 특성요인과 층에 따른 거래가격 차이를 측정오차로 가정하여 필터링함으로써 이를 통제하고 있다.

(반복매매지수) 미국 ...지수에서 사용한 반복매매기법(3단계 추정)을 활용. 전체 지수에서 차지하는 동일주택군별 비중은 단순합산, 거래량, 시가총액 가중 평균방식을 다양하게 적용 가능

(헤도닉모형 적용방안) 동일주택군의 특성정보를 가지고 있는 경우 동일주택군 3000개를 대상으로 헤도닉모형을 적용 가능.

3.5 기존 주택가격지수 추정방법과의 차이

동 연구에서는 이러한 기존 가격지수 산출시 합산방법 외에 주택시장의 대표적인 가격지수 산출방식인 반복매매지수(거래량 또는 세대수 가중회귀), 거래량 또는 시가총액 가중평균 방식으로 합산하여 가격지수를 산출. 이 경우 매기간 모든 대상주택군의 가격이 나오므로 sampling error가 발생하지 않음. 한편 동 방식으로 도출한 개별 주택의 가격을 활용하여 주택의 정보가 확보된다면 헤도닉 방법론으로 적용하는 것도 가능.(추후 논문에서 다룰 예정)

아파트 단지의 평형별 주택을 동의주택의 동일성을 다소 완화하는 대신 측정오차를 상태공간모형과 칼만필터링기법을 활용하여 해결하므로 3000개 아파트 단지의 매월 주택가격을 거래량이 없더라도 추정할 수 있는 방법을 제시. 또한 동 개발 아파트 가격과 해당 아파트 특성정보를 이용하여 헤도닉 모형으로 주택시장의 가격지수를 산출하는 새로운 주택가격지수 산출방법을 제시하였다. 동 주택가격지수는 3천개 아파트 단지의 기초가격을 바탕으로 주택가격지수를 산출하므로 기존 반복매매지수에서 발생했던 sampling Error를 줄일 수 있는 데다 헤도닉 모형의 장점인 ... 등을 가지고 있어 이론적으로 기존 추정방식에 비해 우수한 기법으로 판단된다. 이는 한국 주택시장이 아파트라는 특성을 가지고 있는 점에서 가능 기존지수에 비해 포괄범위 차이

기존 선행연구에서 상태공간모형 적용 논문과의 차이 - 반복매매기법, 헤도닉 기법을 보완하는 정도로 최종 단계에서 사용(합산 이후에 한번) - 본 논문에서는 동일주택군 평균가격 추정시 각각 사용후 합산(합산전에 각 동일주택군별 단계에서 적용)

Chapter 4

실증분석 결과

4.1 동일주택군 단순 중위가격

동일주택 가정 완화로 반복매매거래쌍(정보이용 효율성)이 % 개선 전체거래빈도 대비 반복거래쌍의 빈도는 %의 효율을 보이고 있다. 동일주택군별 거래정보로부터 매월 중위가격을 측정. 동 가격은 층, 인테리어등의 개별 주택특성과 관련된 측정오차를 내재하고 있음. 추정결과 전체 144개월중 최소 50~144개월 내에서 중위가격을 추정 가능

[그림] 개월수 vs 동일주택군(3000개) 분포

동일주택군의 중위가격은 총거래량에 따라 편차의 정도가 다름

[그림] 총 거래량과 편차 [그림] 개별 사례 — 편차가 심함

4.2 동일주택군별 평균가격 추정

아래에서는 추정된 아파트가격의 결과를 특정 아파트 단지에 대해서 살펴보기로 한다. 은 2006~2016년 중 주택 거래량이 1,494건으로 가장 많았던 노원구 중계동의 주공2차 아파트(44.5)(그림 3-a)와 강남지역에서 거래량이 가장 많았던(964건) 대치동 은마 아파트(76.8)(그림 3-b)의 평형별 평균 가격을 상태공간모형으로 추정한 결과이다. 거래량이 풍부한 경우 추정된 아파트가격이 실거래가격과 거의 유사한 움직임을 보이는 것을 확인할 수 있다.

는 강남구 개포동 주공6차 아파트(60.1)의 추정 주택가격과 실거래가격을 보여준다. 해당기간 중(총 132개월) 총 90개월에만 실제 거래가 있었고 나머지 기간에는 거래가 전혀 없음에도 불구하고 추정된 주택가격이 실거래가격과 유사한 움직임을 보이고 있음을 알 수 있다. 는 동작구 신대방동 경남아파트(56)의 추정결과이다. 해당기간 중(총 132개월) 거래실적이 총 77개월에만 있었고 대부분 경우 매월 거래건수가 1건에 불과하여 실거래가격이 거래된 아파트의 특성요인에 따라 크게 변동하는 것을 알 수 있다. 반면 추정된 주택가격은 실거래가격에 비해 비교적 안정적인 움직임을 보이고 있어 Schulz and Werwatz(2004)가 주장한 바와 같이 이 추정방법이 개별 주택의 특성요인을 적절하게 통제하는 것으로 판단된다.

[그림] 평균가격 추정사례 [그림] 추정가격 그래프, 2017년중 상승률 비교

4.3 합산지수 추정 결과

서울지역 아파트가격지수 추정

4.4 기존 아파트가격지수와의 비교

서울지역 아파트가격지수 추정

Chapter 5

결론

본 연구에서는 주택의 동일성을 다소 완화하는 대신 측정오차를 상태공간모형과 칼만필터링기법을 활용하여 해결하므로 3000개 아파트 단지의 매월 주택가격을 거래량이 없더라도 추정할 수 있는 방법을 제시. 또한 동 개발 아파트 가격과 해당 아파트 특성정보를 이용하여 헤도닉 모형으로 주택시장의 가격지수를 산출하는 새로운 주택가격지수 산출방법을 제시하였다. 동 주택가격지수는 3천개 아파트 단지의 기초가격을 바탕으로 주택가격지수를 산출하므로 기존 반복매매지수에서 발생했던 sampling Error를 줄일 수 있는 데다 헤도닉 모형의 장점인 ... 등을 가지고 있어 이론적으로 기존 추정방식에 비해 우수한 기법으로 판단된다. 이는 한국 주택시장이 아파트라는 특성을 가지고 있는 점에서 가능

본 연구는 상태공간모형(칼만필터)을 활용하여 서울지역 개별 아파트의 단지/평형별 아파트 가격을 추정한 후 GARCH(1,1)모형을 적용하여 개별 아파트의 가격 변동성을 추정하는 새로운 방법론을 제시하였다. 이 방식으로 구한 2011~2016년까지의 서울지역 아파트의 가격 변동성은 평균 9.6%로 기존 국토교통부 및 국민은행에서 발표하고 있는 주택가격지수로부터 추정한 가격 변동성에 비해 8%p 가량 더 높은 것으로 나타났다. 또한 2011~2016년 중 개별 아파트의 가격 변동성은 0.6~28.8%로 크게 상이한 모습을 보이고 있으며 이러한 차이는 아파트의 위치, 전용면적, 건축연도 등에 크게 영향을 받는 것을 확인하였다. 본 연구는 상태공간모형을 활용하여 주택가격을 추정했던 기존 선행연구(Schulz and Werwatz, 2004; 박현수·유은영, 2014)과 다음과 같은 점에서 차별화된다. 첫째, 연구방법론 측면에서 선행연구들은 시점별로 거래빈도가 서로 다른(특히 거래빈도가 낮은) 주택시장의 특성과 이에 따른 주택가격지수 산출의 한계를 극복하기 위해 상태공간모형을 이용하여 주택가격지수를 지역별로 산출한 반면, 본 연구에서는 개별 아파트 단지의 동일 평형 아파트를 동일한 아파트로 가정하여 분석기간 동안 일정한 수준의 거래(최소한 70개월 이상)가 있는 모든 동일 아파트에 대해 아파트 가격을 추정하였다. 둘째, 상태공간모형(칼만필터링)을 이용한 목적과 범위도 상이하다. 선행연구에서는 서로 다른 주택의 거래가격 또는 반복매매상에서 각 주택의 특성요인을 배제한 주택가격의 공통요인(common price component)을 추정하기 위한 목적에서 상태공간모형을 이용하였지만, 본 연구에서는 각 아파트 단지내 동일평형 아파트의 거래가격에서 방향, 층과 내부구성 등의 차이를 배제한 평균 거래가격을 필터링할 목적에서 사용하였다. 결과적으로 본 연구에서 제시한 결과는 통제(필터링)하고자 하는 요인이 제한적이며 기초데이터의 동일성이 더 높다고 할 수 있다. 셋째, 본 연구에서는 상태공간모형을 통해 추정된 개별 아파트가격에 기초하여 개별 아파트 가격의 변동성을 추가적으로 추정하였다. 부동산 투자는 개별적인 투자대상에 기초하고 아파트가격의 변동성이 투자의 한 가지 위험요인임을 고려하면 본 연구방법을 이용하여 추정되는 개별 아파트가격의 변동성은 중요한 투자지표가 될 수 있을 것으로 생각된다. 또한 전세계약과 같이 옵션이 내재된 계약의 경우 개별 아파트의 전세계약별로 옵션가치를 측정하는데 유용하게 활용될 수 있을 것이다. 그러나 본 연구는 다음과 같은 점에서 한계가 있다. 첫째, 주택가격을 추정하기 위해 최소한의 시계열 자료가 필요하므로 최근 5년 내에 신축된 아파트나 거래량이 거의 없는 아파트의 경우 신뢰성 있는 변동성을 추정하기 어렵다. 둘째, 본 연구에서는 주택가격의 변동성을 추정하기 위한 새로운 방법론을 제시하는데 목적으로 두었기 때문에 가장 대표적인 변동성 추정방식인 GARCH(1,1) 모델을 사용하였으나 부동산 시장의 경우 가격 상승과 하락기에 변동성의 움직임이 다르게 나타날 수 있으므로 추정결과의 신뢰도를 제고하기 위해서는 EGARCH 등 다양한 변동성 추정방법에 대한 실증분석이 필요하다. 셋째, 본 연구에서는 개별 아파트의 가격변동성을 추정된 평형별 아파트가격과 평형별 세대수를 이용한 가중치를 이용하여 지역별 가격변동성을 집계하였다. 그러나 가중치를 구하는 방법은 여러 가지가 있을 수 있으므로 어떤 방식의 가중치 계산이 적절한지에 대한 보다 구체적인 연구도 필요할 것으로 생각된다. 마지막으로 본 연구에서는 가격 변동성의 지역별, 전용면적별, 건축연도별 차이와 주택가격 변동률 및 거래량과의 관계를 단순 비교하였는데 이에 대한 보다 심도있는 분석이 요구된다.

Appendix A

칼만필터링 A

경제분석에서는 직접 관측이 불가능한 개념적 변수를 관측 가능한 관련 자료로부터 추출하고자 하는 경우 상태공간모형(state space model)을 활용한다. 상태공간모형은 비관측변수인 상태변수(x_k)와 관측변수(z_k)간의 관계를 표시하는 관측방정식과 비관측 상태변수의 전이(transition)과정을 나타내는 상태방정식으로 구성된다(Harvey, 2001). 본 연구에서는 개별 아파트단지의 평형별 아파트의 월별 평균 가격수준을 직접 관측이 불가능한 변수라고 가정한다. 한편 관측가능한 자료인 해당(특정 단지, 특정 평형) 아파트의 월별 평균 실거래 가격은 해당 아파트의 평균적인 가격수준과 실제 거래된 아파트의 특성요인(남향여부, 층수, 인테리어 등) 및 오차항으로 구성된다고 가정한다.

상태방정식 :

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \quad (\text{A.1})$$

관측방정식 :

$$z_k = Hx_k + v_k \quad (\text{A.2})$$

$$w_k \sim N(0, Q)$$

$$v_k \sim N(0, R)$$

식 (1)와 식 (2)에서 A, B, H 는 각각 상태변수내 전이과정, 추가적인 독립변수와 상태변수간 관계, 관측변수와 상태변수간 관계를 나타내는 변수(행렬)이다. w_k 와 v_k 는 각각 process 및 measurement noise를 의미하며 서로 독립적이며 정규분포를 따른다고 가정한다¹. 한편 \hat{x}_k^- 을 k 시점 이전의 모든 정보를 바탕으로 k 시점의 상태변수값을 추정한 사전추정치(priori state), \hat{x}_k^- 를 관측변수 측정값(z_k)를 포함하여 k 시점의 상태변수값을 추정한 사후추정치(posteriori state)라고 할 때 사전·사후 추정오차(e_k^-, e_k)와 관련 공분산(P_k^-, P_k)은 아래와 같이 정의된다.

사전 추정오차 및 공분산 :

사후 추정오차 및 공분산 :

상태공간모형의 추정은 주로 칼만필터 알고리즘(Kalman, 1960)을 활용한다². 칼만필터는 상태변수의 사후추정치(\hat{x}_k)를 사전추정치(\hat{x}_k^-)와 관측변수의 측정오차($Z_k - H\hat{x}_k^-$, measurement innovation or residual)간 선형결합으로 표현하는 아래의 식 (5)과 같은 최적의 칼만방정식을 도출함으로써 상태변수의 사후추정치(\hat{x}_k)를 추정한다.

¹연구목적에 따라 A, B, H 변수(행렬)와 w, v 의 조건부 분산인 Q 와 R 이 시간에 따라 변한다고 가정하기도 한다.

²칼만필터는 과거의 측정데이터와 새로운 측정데이터를 사용하여 데이터에 포함된 노이즈를 제거시켜 보다 개선된 새로운 추정을 하는 알고리즘으로 현재 모든 스마트폰과 네이게이션 기기 등 과학분야에서 널리 활용되고 있다.

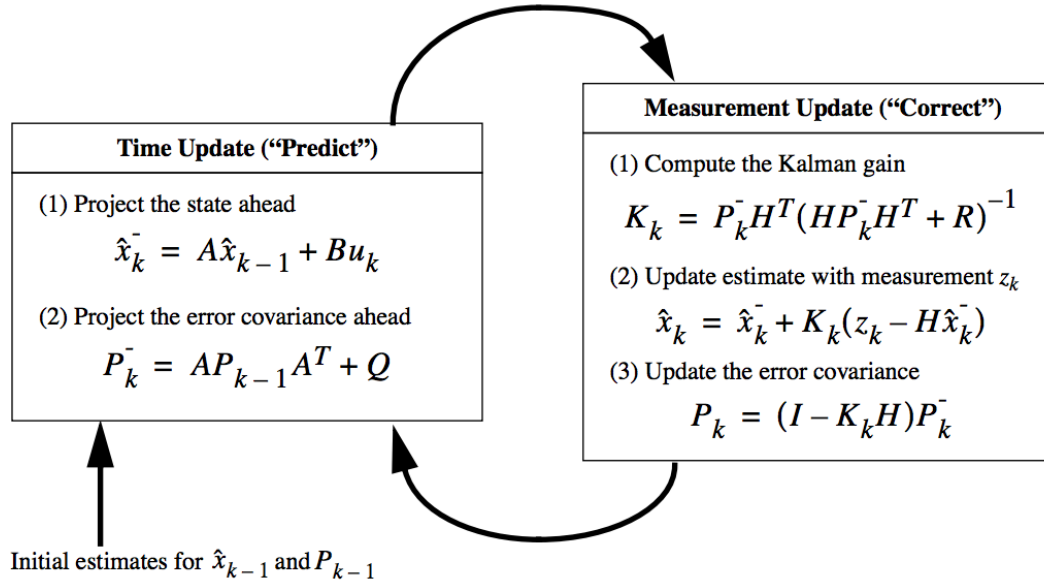


Figure A.1: 칼만필터 알고리즘 추정방법 (Welch and Bishop, 2001)

여기서 칼만방정식은 사후 추정오차의 공분산($P_k = E(e_k e_k^T)$)을 최소화하는 최적의 K 값을 도출함으로써 구할 수 있으며 이를 Kalman gain이라고 한다. Maybeck (1979), Jacobs (1993), ? 칼만방정식의 최적 K 값과 예측오차공분산(P_k)을 다음과 같이 도출하였다.

결과적으로 사후추정치들은 사전추정치의 예측과정(time update)과 새로운 관측치를 반영하는 수정과정(measurement update)을 반복하면서 추정된다. 여기서 Kalman gain은 사후추정치 계산할 때 사전추정치와 새로운 관측변수를 반영한 측정오차간의 최적 가중치를 의미한다. 즉 칼만필터는 사전추정치와 새로운 관측변수의 예측오차를 사후적으로 비교하여 더 정확한 변수에 더 높은 가중치를 계속 부여함으로써 사후예측결과를 계속 개선하는 알고리즘이라고 할 수 있다. 아래 은 칼만필터 알고리즘의 추정방법을 보여준다³.

```
knitr::include_graphics("images/figure20.png")
```

³상태공간모형과 칼만필터링 기법에 대한 보다 자세한 내용은 Harvey (2001), Welch and Bishop (2001) 등을 참조할 수 있다.

국문요약

개별 주택가격과 변동성 추정에 관한 연구 - 서울지역 아파트 실거래에 기초한 상태공간모형과 GARCH모형 적용 -

본 연구에서는 Holmes(2009)의 KalmanEM(R package)를 활용하여 상태공간모형으로 서울지역의 개별 아파트단지의 전용면적별 아파트가격을 추정하였다. 그리고 추정된 아파트가격을 이용하여 GARCH(1,1) 모형을 적용하여 개별 아파트가격의 변동성을 추정하였다. 이러한 방법을 적용함으로써 개별 아파트의 특성요인을 효과적으로 통제할 수 있으며 기존의 주택가격지수로부터 주택가격의 변동성을 추정할 경우 변동성이 낮게 추정되는 문제를 극복할 수 있었다. 실증분석 결과, 2010~2016년 기간 동안 서울지역 주택가격의 월평균 변동성(연율)은 9.6%로 국토교통부, 국민은행의 주택가격지수로부터 추정한 변동성(각각 1.2%, 1.4%)에 비해 약 8%p 가량 높게 추정되었다. 지역별로는 강남구, 양천구 등 일부지역의 가격 변동성이 높았으며 기타 지역과는 5%p 가량 차이가 지속되고 있는 것으로 나타났다. 또한 건축연도가 오래되고 전용면적이 작은 아파트의 가격 변동성이 높게 추정되었다. 아파트가격 변동성과 아파트 거래량 및 아파트가격 변화율간의 관계는 2010년 이후 크게 약화되었으며, 변동성이 두 변수를 2~3개월 정도 후행하는 것으로 분석되었다.

Bibliography

Harvey, A. C. (2001). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge, UK.

Jacobs, O. (1993). Introduction to Control Theory. University Press, 2nd edition.

Maybeck, P. S. (1979). Stochastic Models, Estimation, and Control. Academic Press.

Welch, G. and Bishop, G. (2001). An Introduction to the Kalman Filter. ACM, Inc.

이용만·이상한 (2008). 국민은행 주택가격지수의 평활화 현상에 관한 연구. 주택연구, 16(4):27-47.