

# 아파트시장예측을 위한 신경망분석 적용가능성에 대한 연구

## A Study on the Applicability of Neural Network Model for Prediction of the Apartment Market

남 영 우\* · 이 정 민\*

Nam, Young-Woo · Lee, Jeong-Min

### 요 약

부동산분야에서 전통적인 예측방법과 비교하여 보다 예측력을 높일 수 있는 방법을 찾으려 한다. 이에 앞서 신경망 모형의 적용가능성을 살펴보고, 기존의 연구를 토대로 한 신경망 이론의 정의, 구조, 장단점 등을 살펴본다. 구체적인 적용가능성을 확인하기 위하여 동일 데이터로 회귀분석과 신경망분석을 통한 모형을 구축하고, 예측정확도 측면에서 신경망모형의 적용가능성을 검토한다. 부동산학에서 기존에 회귀분석에 치우쳐 있던 연구방법을 신경망분석까지 확장하고, 특히 예측정확도 측면에서 우수성이 검증되고 있는 신경망모형에 대한 연구를 활성화 하고자 하는데 본 연구의 목적이 있다. 연구방법으로는 분양가격에 영향을 주는 거시경제변수를 모형화 한다. 그 모형설정 후 회귀분석과 신경망분석으로 결과를 비교하여 보다 예측 정확도가 높은 것을 찾는다. 그 결과 신경망모형의 예측정확도가 상당히 높게 나타났다.

키워드: 부동산시장, 신경망모형, 회귀모형

## 1. 서론

### 1.1 연구 배경

IMF이후 국내 경제가 많은 변화를 겪으면서 부동산실물 시장 뿐 아니라 학문적 측면에서도 부동산 시장을 분석하고 예측하기 위한 다양한 연구의 필요성이 대두되었다. 따라서 시장의 수요와 공급 및 가격변동 등과 같은 중요한 현상들에 대한 예측을 위한 다양한 연구가 전개되었고, 이러한 연구결과의 축적은 부동산에 대한 이해를 넓히고 과학적인 접근의 필요성을 인식하는데 크게 기여하였다. 특히, 주택시장을 중심으로 한 가격요인의 분석 및 예측에 대한 연구가 활발히 진행되었으며, 방법론으로는 시계열을 반영한 VAR모형과 특성함수(hedonic function)분석을 중심으로 한 회귀모형이 주로 사용되었다.

특히, 아파트시장에서는 정보제공업체를 통하여 매매가 및 아파트 특성에 대한 자료에 대한 접근이 용이하게 되면서 회귀모형을 중심으로 한 연구가 활발하게 이루어 졌다. 회귀모형을 통한 연구 성과가 축적됨과 함께 회귀모형을 보완할 수 있는 새로

운 방법론을 통한 연구에 대한 필요성 또한 증가하고 있다. 이는 회귀모형이 가지는 통계적 장점에도 불구하고 회귀모형은 등분산성, 독립성, 정규성 등의 가정과 다중공선성의 문제, 특이값에 민감하게 반응하는 문제점 등이 있으며, 모수의 비선형성 고려 문제와 최적회귀식을 선택하였는지에 대한 의문점 등이 제기될 수 있기 때문이다.(김태훈외 1인 2004) 따라서, 회귀모형의 문제점을 극복하고 회귀모형과 상호보완적인 관계를 가질 수 있는 새로운 모형에 대한 관심이 높아지고 있다.

이렇게 회귀분석 외에 새로운 방법론의 필요성이 대두되면서 최근 신경망이론에 대한 관심이 고조되고 있다. 신경망은 입력변수가 불완전하고 변동 폭이 넓은 경우에도 해석이 가능하고 데이터 수가 적거나 불규칙한 경우에도 적용이 가능하기 때문이다. 신경망이론은 특히 예측분야에서 유리한 조건을 갖추고 있다. 왜냐하면 기존의 예측모형들이 엄격한 통계적 가정 하에 기초하여 만들어져 상황의 변화나 복잡한 환경변수들의 영향을 고려하지 않았기 때문이다.(이건창 2000)

신경망모형은 초기 인공지능을 개발하기 위하여 공학 분야에서 발전하였으나, 최근 경영학에서 많이 사용되고 있다. 특히, 고객만족도 및 우수고객을 선별하는 등 마케팅 분야에서 많이 활용되고 있으며, 주식시장에서의 주가예측이나 기업의 부도가능성예측 등 재무 분야에서도 많이 적용되고 있다. 따라서 부동산 분야에서도 그 개념에 대한 이해 및 적용가능성에 대한 연구

\* 일반회원, 건국대학교 부동산학과 박사과정, ywnam72@paran.com

\*\* 일반회원, 경희사이버대학교 자산관리학과 전임강사, 부동산학박사(교신저자), ljm001@khcu.ac.kr

의 필요성이 대두되고 있다.

지금까지 부동산 분야에서 신경망모형을 적용한 연구는 신경망모형을 적용한 토지평가시스템구축에 대한 연구(최병남 1994)와 건교부의 개별공시지가 산정에 신경망을 응용한 연구(정화미외 2인 2001), 아파트의 특성변수를 이용하여 회귀모형과 신경망모형을 구축하고 이를 비교한 연구(김태훈외 1인 2004)가 있었다. 각각의 연구가 부동산 분야에서 아직까지 일반화되지 않은 방법론인 신경망모형을 소개하고 적용하였다는 데는 의미가 있으나, 각 연구간 연계성이 없이 특정 사안에 적용하는데 그쳐 신경망을 통한 부동산시장에 대한 연구가 지속적으로 발전하는 데는 한계가 있었다. 따라서 신경망이론을 통한 연구의 발전을 위해서 이론적 배경에 대한 소개와 부동산 분야에 대한 적용가능성을 제시하는 연구가 필요한 시점이 되었다.

## 1.2 연구목적

본 연구에서는 먼저 신경망 모형의 이론적 배경 및 개념을 살펴보고, 부동산분야에서의 적용가능성을 살펴보고자 한다. 따라서 기존의 연구를 토대로 한 신경망 이론의 정의, 구조, 장단점 등을 살펴보고, 구체적인 적용가능성을 확인하기 위하여 동일 데이터로 회귀분석과 신경망분석을 통한 모형을 구축하고, 예측 정확도 측면에서 신경망모형의 적용가능성을 검토하고자 한다.

이러한 연구를 통해서, 부동산학에서 기존에 회귀분석에 치우쳐 있던 연구방법을 신경망분석까지 확장하고, 특히 예측정확도 측면에서 우수성이 검증되고 있는 신경망모형에 대한 연구를 활성화 하고자 하는데 본 연구의 목적이 있다.

## 2. 신경망 모형의 이론적 고찰

### 2.1 인공신경망 모형의 정의

#### 2.1.1 생물학적 신경망

인공신경망은 결국 생물학적 신경망의 특성을 반영하여 모형화 한 것이므로 인공신경망에 대한 이해를 위해서는 생물학적 신경망에 대한 이해가 선행되어야 한다.

인간의 두뇌는 뉴런(neuron)이라고 불리는 세포로 구성되어 있다. 인간 두뇌에 있는 뉴런의 수는 1,000억 개에 이르고 1,000가지 이상의 다른 뉴런이 있는 것으로 알려지고 있다. 뉴런은 네트워크(network)이라고 불리는 집단들에서 기능을 하고, 각 집단은 수천 개의 잘 연결된 뉴런들로 구성되어 있다. 따라서 두뇌는 신경망(neural network)의 집합체로 간주될 수 있다.(김성무 2001)

뉴런은 신호를 생성하고 전달하여서 체세포에 전달하는 역할을 하며 이 과정에서 두뇌가 학습능력을 가지게 하는 기능을 한다. 이러한 뉴런을 통한 정보전달체계가 모형화 한 것이 인공신경망모형이 된다.

#### 2.1.2 인공신경망의 역사

인공신경망 연구의 시작은 1943년 McCulloch와 Pitts에 의해 시도된 뉴런(Neuron)의 모형화이다. 이들이 제시한 신경망 모델은 신경세포와 그것들의 연결로 이루어져 있으며, 신경세포의 출력값은 1 또는 0인 이진 형태의 동작을 한다. 이 모델은 뉴런을 하나의 처리단위로서 계산 기능을 중요시하였으나 나머지 학습기능에 대해서는 도외시 하였다.

1949년 캐나다의 심리학자 헵(Hebb, D.)은 시냅스(Synapse)라고 알려진 연결강도조정을 위한 학습규칙을 최초로 기술하였다. 학습규칙의 기본원리는 신경세포들 간의 연결가중치를 조절함으로써 신경망을 학습시킨다는 것이었다. 이를 통해 인공신경망이 학습을 수행한다는 사실이 알려지게 되었다.

1950년대와 60년대에는 생물학 및 심리학적 이론을 활용한 인공신경망을 생성하려는 연구가 활발히 이루어졌다. 1957년에 Rosenblatt는 Hebb의 규칙에 이어 두 번째 학습규칙을 개발하였다. 델타 규칙(delta rule)이라고 불리는 이 규칙은 주어진 입력에 대한 신경망의 실제 출력과 우리가 원하는 출력간의 차이를 이용하여 학습을 수행한다. Rosenblatt는 이 델타규칙을 적용하여, 학습시키고자 하는 입력과 그 입력에 대응하는 출력을 제시하는 것만으로도 학습을 수행할 수 있는 퍼셉트론(perceptron)이라는 신경망을 만들어 냈다.(김성무 2001)

그러나 1969년 Minsky와 Papert는 단일계층 퍼셉트론은 XOR함수와 같은 단순한 비선형 분리분제도 해결할 수 없다는 것을 밝혀내어 인공신경망의 한계를 지적하였고, 이 후 20년 동안 신경망에 대한 연구는 침체의 길을 걷게 되어 암흑기를 맞이하게 되었다.

80년대 들어 인공신경망 연구의 암흑기를 벗어나게 하는 두 가지 사건이 발생하였다.(정유석 2003) 첫 번째 사건은 1982년 Hopfield가 인공신경망의 행위를 분석하고 학습시키는 것이 가능하다는 것을 증명한 것이며, 두 번째 사건은 Werbos 등에 의해서 다층퍼셉트론을 학습시킬 수 있는 역전파 학습알고리즘이 발견된 것이다. 이를 통해서 이후 신경망에 대한 연구가 활발하게 진행되어 인공신경망의 응용분야도 패턴분류나 연상기억장치의 범주에서 벗어나, 최적화, 로봇제어, 문자인식, 음성인식, 신호처리 등으로 확대되었으며, 최근에는 경영학 등 사회과학 분야에서도 많이 활용되고 있다.(정유석 2003)

2.1.3 인공신경망 모형의 정의

인공신경망 모형(Neural Network)은 생물학적 신경망을 컴퓨터를 이용하여 모형화 하려는 노력에서 나타난 것으로 인간 두뇌의 신경망을 흉내 내어 자신이 가진 데이터로부터 반복적인 훈련 과정을 거쳐 데이터에 숨어있는 패턴을 찾아내는 모델링(modeling)기법이다.

신경망이론은 엄청난 계산 속도를 필요로 하는 문제영역에 있어서 기존의 계산방법에 비해 훨씬 높은 가능성을 가지고 있는데 이는 기존의 해결방식이 기본적으로 순차적인 명령문으로 구성된 프로그램을 수행하는데 반해, 신경망이론은 많은 처리단위(processing element)혹은 노드(node)로 이루어진 병렬 네트워크(massively parallel nets)를 이용하여 여러 개의 가능성을 동시에 탐구하는 방식이기 때문이다.

인경신경망의 노드구조를 단순화하여 나타내면 <그림1>과 같다.

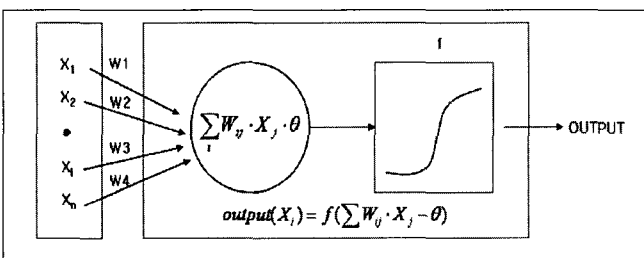


그림1. 인공신경망의 노드 구조

하나의 노드는 다른 노드들로부터 입력을 받지만 자신은 Xi 라고 표기된 단 하나의 출력값만을 생성하고 이 값을 또다시 연결된 다른 노드에 전달한다. 다른 노드들로부터 전달받은 X1, X2, X3, ..., Xj, ..., Xn 등 n개의 입력값은 각각의 연결선에 가해지는 i와 j번째 노드사이의 연결가중치 Wij와 곱해진 후, 이들의 가중치 합이 어느 임계치 θ 이상이 되면 그 노드는 수행되어 아래와 같이 순입력값을 계산한 후 비선형함수 f를 활성화시켜 노드의 최종적인 출력값을 얻게 된다. (정유석 2003)

순입력값 :  $net_i = \sum W_{ij}X_j$

최종출력값 :  $x_i = f_i(net_i)$

이때 순입력값을 출력값으로 변환시키는 함수 f를 전이함수라고 한다.

즉, 전이함수란 처리단위의 처리결과를 출력으로 변환시키는 함수를 말하는데 신경망에서 사용하는 가장 단순한 전이함수로는 가중 합산된 값을 그대로 사용하는 선형함수(linear function)가 있으며, <그림2>의 (a)에서 보는 것처럼 특정한 임계값을 기준으로 0 또는 1과 같은 값을 취하는 임계논리(threshold logic)와, 0이하에서는 0의 값을 갖고 1이상에서는 1

의 값을 가지며 0과 1사이에서는 선형인 램핑함수, 그리고 S자 모양의 시그모이드함수(sigmoid function) 등이 이용된다.(김성무 2001)

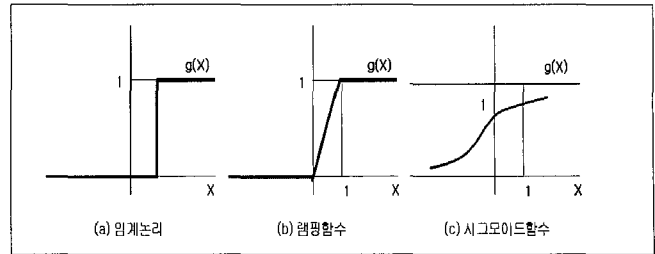


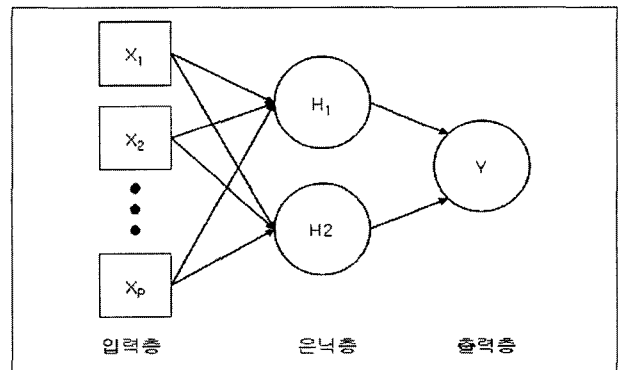
그림2. 전이함수의 유형

2.2 신경망 모형의 구조

앞서 인공신경망 연구의 역사에서 나타난 바와 같이 최근 사용되고 있는 신경망모형은 일반적으로 다층퍼셉트론의 구조와 다층퍼셉트론의 학습방법인 역전파 학습방법을 통해서 구축된다.

2.2.1 다층퍼셉트론

다층 퍼셉트론은 입력층과 출력층 사이에 하나이상의 은닉층이 존재하는 모형으로 <그림3>과 같은 계층구조를 갖는다. 입력층은 외부로부터 신호를 받아서 은닉층으로 분산시켜 주는 기능을 한다. 은닉층과 출력층에서는 학습규칙에 따라 입력층과 원하는 출력을 이용하여 학습하고 인식하는 역할을 한다. 각 뉴런은 외부로부터의 입력에 가중치를 곱하고 이를 모두 합한 다음, 활성화 함수를 통하여 다음 뉴런으로 신호를 전달한다.(정화미 외 2명 2001)



$H1 = f_1(b_1 + w_{11}X_1 + w_{21}X_2 + \dots + w_{p1}X_p)$   
 $H2 = f_2(b_2 + w_{12}X_1 + w_{22}X_2 + \dots + w_{p2}X_p)$   
 $y = g(b_0 + W_{10}H_1 + W_{20}X_2H_2)$

그림3. 다중 퍼셉트론 구조

위의 수식에서  $f_1$ ,  $f_2$ 와  $g$ 를 각각 활성화함수와 출력함수라고 한다. 활성화함수는 로지스틱 회귀모형에서 연결함수의 역함수와 유사한 의미를 가지며 가장 보편적으로 로지스틱함수가 사용된다.

### 2.2.2 역전파(Back Propagation) 알고리즘

역전파알고리즘의 기본원리는 입력층의 각 노드에 입력자료가 주어지면 이 값은 각 노드에서 연결강도와 결합하여 다음 층에 전달되고 최후의 출력층에서 최종값을 출력하게 된다. 인공신경망에 의해서 계산된 이 출력값과 목표값을 비교하여 그 차이를 줄여나가는 방향으로 상위층에서 역전파하여 하위층에서는 이를 근거로 다시 자기층의 연결강도를 조절해 나간다.(김대수 1993)

역전파 학습알고리즘의 학습과정은 <그림4>와 같다.

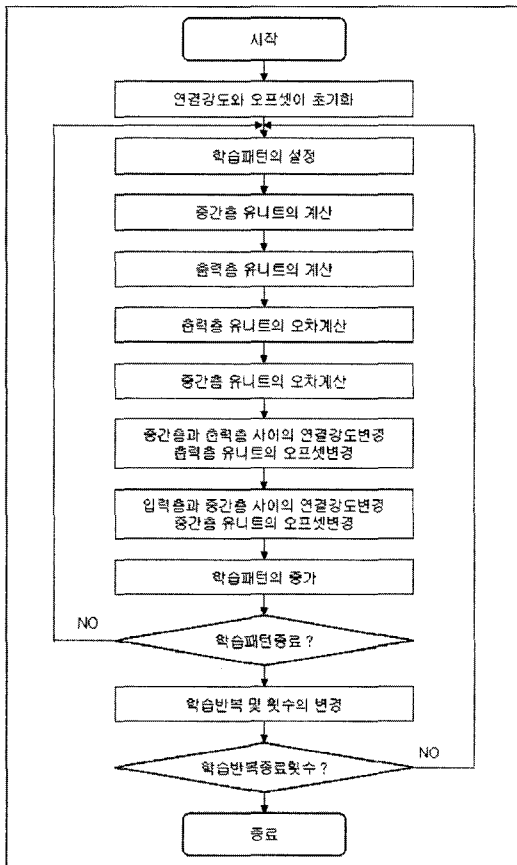


그림4. 역전파 학습 알고리즘의 학습과정

### 2.3 인공신경망모형의 장·단점

인공신경망은 다른 분석모형에 비해서 다음과 같은 장, 단점을 가지고 있다.

#### 2.3.1 인공신경망의 장점

첫째, 인공신경망은 병렬분산처리(parallel processing)가 가능하다. 즉, 인간의 두뇌처럼 여러 개의 뉴런이 입력값에 대해서 동시에 연산을 수행함으로써 짧은 시간에 패턴인식을 할 수 있다.

둘째, 일부분의 결합에 대해서 전체 네트워크가 쉽게 손상되지 않는다. 즉, 처리가 많기 때문에 몇 개의 노드나 연결이 가진 결합이 시스템전체에 결합을 초래할 가능성이 낮다.

셋째, 입력치로부터 반복적인 실험과 훈련을 통하여 가중치를 알아내어 규칙성을 발견함으로써 새로운 입력값에 대해서 합리적인 반응을 생성할 수 있다. 따라서 예측을 위한 모형설정에 유용성이 높은 모형이다.

#### 2.3.2 인공신경망모형의 단점

인공신경망모형의 단점으로 지적되고 있는 내용의 근본 원인은 이론적 근거가 약하고 산출과정에 대한 명확한 제시가 어렵다는 점이다. 인공신경망 모형의 단점은 다음과 같다.

첫째, 인공신경망모형은 입력변수와 출력변수와의 상호관계를 이해하기가 어렵다. 따라서 인공신경망 모형은 결과값이 결과의 도출과정보다 중요한 경우 사용될 수 있는 방법이다. 인공신경망 모델의 내부가중치 들은 입력층, 은닉층, 출력층 내에서 함께 얽혀있어 변수간의 상관관계를 명확하게 제시하지 못하게 된다. 다만, 판별력 분석을 통하여 여러 개의 값 중에서 어느 것이 중요인지 구별할 수는 있다.

둘째, 모든 입력값과 출력값이 0과 1사이의 범위에서 정의되어야 한다. 인공신경망에서 사용되는 입력값은 특정 범위로 변환되어야 하기 때문에 입력자료에 대한 변환작업 등이 필요한데, 이 과정에서 변환방법의 선택이 인공신경망의 결과값에 영향을 미치게 된다.

셋째, 은닉층의 최적수를 결정할 수 있는 이론적 토대가 없다. 즉, 인공신경망의 은닉층과 은닉마디 수는 확실한 이론적 배경보다는 모형의 시행착오를 통해 만들어지므로, 은닉층 수의 선택에 대한 이론적 토대가 없다는 문제가 나타난다.

### 2.4 인공신경망의 적용범위

초기 인공지능의 구현을 위해서 공학분야에서 연구되었던 인공신경망 모형은 최근 경영 및 재무분야에 까지 활용되고 있다. 특히, 기업도산예측, 주식투자시점결정, 주가지수 선물거래 응용, 기업공개 시 IPO가격 결정 등 재무분야에서 매우 인기 있는 도구로 활용되어지고 있다. 투자분석문제에서 데이터와 정보의 대부분은 계량적 데이터이며 종종 많은 이상치 때문에 분석이

곤란한 때도 있다. 따라서 선형성과 정규확률분포와 같은 가정에 바탕을 둔 통계분석들의 대부분은 투자분석에 있어서 제한적으로 사용될 수밖에 없었다. 또한 인공지능망은 같은 인공지능의 분야로 전문가의 도움 없이 데이터를 통해 지식을 도출해내는 기계적 훈련의 장점을 가지고 있다. 위와 같은 이유로 인공지능망 모델은 재무응용분야에서 주목을 받고 있다. (서영호 외 1998)

또한, 일반 기업체에서는 고객의 거래내역과 인구통계학적 변수를 통해서 고객의 만족도, 기업체에 대한 충성도를 분석하고, 신규고객의 향후 수익성을 예측하는데 유용하게 사용하고 있다.

재무 및 경영분야에서 인공지능망의 유용성이 확인되고 그 적용범위가 넓어지는 것은 부동산부문의 적용가능성에 청사진을 제시하는 것으로 볼 수 있다. 아직까지 부동산 분야도 분석에 필요한 데이터의 구축이 불완전하기 때문에 전통적인 통계분석을 적용하는데 한계가 있는 경우가 많이 있고, 회귀분석 등을 통한 예측모형을 보완하기 위해서도 유용하게 사용가능할 것으로 예상된다. 또한, 정부정책 수립 시에도 다양한 변수의 적용을 통한 새로운 정책도입 시 시장환경변화의 예측이나 공시지가 및 보상가격의 적정성 검증을 위해서도 적용가능 할 것으로 예상된다.

### 3. 회귀분석과 신경망분석 결과의 비교

#### 3.1 분석개요

##### 3.1.1 분석의 의의 및 범위

부동산시장에 대한 신경망 모형의 적용가능성을 파악하기 위해서 본 연구에서 다루는 부동산 시장모형은 주택시장 및 토지시장을 중심으로 한 부동산과 주요 거시경제 지표를 포함하는 부동산계량모형이다.

부동산가격의 예측은 크게 향, 평수, 위치 등 부동산자체의 내부적 특성을 통해서 분석하는 방법과 물가상승률, 지가변동률 등의 거시경제변수를 통해서 예측하는 방법이 있다. 최근 아파트의 내부적 특성변수들을 활용한 연구에서 신경망분석을 통한 가격예측이 회귀분석보다 더 우수한 결과가 제시되었다. (김태훈 외 1인 2004)

본 연구에서는 거시경제변수를 통한 아파트 분양가예측에 신경망모형을 적용함으로써 신경망모형의 적용가능성을 확대하고, 훈련데이터와 검증데이터를 분리하여 예측정확도를 산출하여 좀 더 정확한 모형을 개발하고자 한다. 분석을 위한 자료는 통계자료의 이용가능성과 한국 경제의 구조변화를 감안해 1998년부터 2004년1월까지의 서울지역 동시분양아파트 가격과 거시경제 지표를 분석의 대상으로 하고 있다.

분석을 위하여 1998년부터 2001년 데이터를 통해 거시경제변수를 독립변수로, 아파트 분양가를 종속변수로 하는 회귀분석을 실행하고, 동일 변수를 통하여 신경망분석을 실행한 후, 2002년 1월부터 2004년 1월의 데이터를 통해 추정결과를 산출하고, 각각의 결과를 비교평가하여 신경망 분석의 적용가능성을 제시하고자 한다.

##### 3.1.2 변수선정의 유의점

부동산시장모형은 부동산에 직간접으로 영향을 미치는 주요 변수를 포함해 작성되었다. 그렇지만 한국경제에서 차지하는 비중이 계속 높아지고 있는 해외부분의 수출, 수입 등이 모형에서 제외 되어 있어서, 부동산 시장 모형은 몇 가지 제약을 가지고 있다.

부동산시장모형에 내제된 문제점의 하나는 외환위기 당시의 경제왜곡현상이 모형 전체에 영향을 주었다는 점이다. 또한 연구본래의 목적상 거시경제변수들만 설명변수에 포함이 되어야 마땅할 것이지만, 평당 분양가의 왜곡현상 즉, 강남 서초 송파구의 동시분양 물량이 일시적으로 많을 경우, 모형의 적합성에 끼칠 해악을 우려하여 설명변수에 지역이라는 변수를 DUMMY변수로 추가하였으며, 평형별 평당 가격의 가중평균 개념을 적용하기 곤란하여, 평형대가 높으면 평당 분양가격은 비싸지는 원리를 반영하기위해 평형변수를 설명변수로 추가하였다.

따라서 회귀분석에서는 거시경제변수, 지역더미변수, 평형 등이 독립변수가 되어 종속변수인 분양가를 예측하는 모형을 산출하며, 신경망분석에서는 회귀분석의 독립변수가 input 데이터가 되며, 분양가를 output데이터로 하여 모형을 생성한다.

##### 3.1.3 변수의 설정

본 연구에서는 주택가격의 예측을 위하여 기준에 거시경제변수를 통한 부동산시장예측 위해서 사용된 변수(서승환, 1994)를

표1. 분석사용변수

변수유형	변수명	주요내용	단위	
종속변수	PRICE	아파트 평당분양가격	만원	
독립변수	거시 경제변수	CPI	소비자 물가	2000년=100
		M2	총통화량	
		GDP	국민총생산	10억원
		OIL	국제원유도입단가	\$/배럴
		EXCHANGE	환율	원
		COST	건축원자재 도입단가	
		LAND	지가변동률지수	1997년1월 =100
지역 특성변수	AREA	지역 DUMMY		
	PY	평형		

중심으로 분석하였다. 소비자 물가(CPI)와 총통화량(M2), 환율(EXCHANGE), 원자재 도입단가(COST), 지가변동률(LAND), 그리고 국제원유도입단가(OIL) 등의 거시경제변수와 모형의 적합성을 위해 평형이라는 변수와 더미변수로 지역을 추가시켰다.

주택가격에 영향을 미치는 설명변수들의 추정선형함수 식을 도출하기 전에 우선 이러한 변수들의 상관관계를 분석하여 모델 설정을 위한 각 변수의 적합도를 규명하였다. 개별 설명변수 중 CPI는 소비자 물가지수로서, 재정경제원에서 발표하는 연도별 분기별 소비자 물가지수를 사용하였으며, M2는 총통화량으로 통계청 주요 경제지표상의 연도별 총통화액을 의미한다. GDP는 국내 총생산 금액을 의미하며, OIL은 국제원유 도입단가를 의미하며, EXCHANGE는 환율을 의미하며, COST는 통계청에서 원자재 도입단가를 의미한다. LAND는 서울시 분기별 지가변동률(서울시 동시분양만을 고려)을 1997년 1월을 기준으로 조정된 변수이며, AREA는 더미 변수로서 서초구, 강남구, 송파구는 1, 그 외 지역은 0이며 PY는 평형별 평당가격을 고려하기 위한 신규공급주택의 평형을 의미한다.

이러한 영향변수들에 대한 시계열 자료의 시간적 범위는 1998년 1월부터 2004년 6월까지 7개년도로 한정하였으며, 분석의 목적이 다중회귀분석을 통한 예측의 정확도와 신경망 분석을 통한 예측정확도의 비교이므로, 예측정확도의 기준은 모형을 통한 예측값과 실제값과의 오차를 이용한 MSE를 사용하였다. 그리고 기존연구에서 모형구축에 사용된 데이터를 통한 예측의 정확성과 모형구축에 사용되지 않은 변수의 예측정확도에 차이가 나타났으므로(이현석 2004), 1998년 1월부터 2001년 12월 까지의 데이터인 2,634개의 분양데이터는 모형구축을 위한 데이터로 사용하였으며, 2002년 이후 데이터는 모형의 적합도 검증을 위하여 사용하였다.

3.2 회귀분석 결과

3.2.1 다중 회귀분석 개념

다중회귀분석이란 종속변수의 변화를 설명하기 위하여 두 개 이상의 독립변수가 사용되는 선형회귀모형을 말하며, 실제 자료를 이용한 모형적합의 경우에는 대부분이 중회귀모형을 이용한 모형적합이 이용된다.

3.2.2 다중회귀분석의 필요성

단순회귀모형은 하나의 독립변수와 종속변수로 구성된다. 그러나 실제로 단일 요인에 의해서 결정되는 현상은 매우 드물다. 대부분의 인과관계구조는 여러 요인들로 복잡하게 얽혀있기 마련이며, 따라서 다수의 독립변수를 모형에 포함시키는 다중회귀

분석이 불가피해진다. 주택의 가격 등을 다양한 특성변수를 통해 예측하는 모형을 HEDONIC PRICE FUNCTION이라고 하며, 이 모형은 주택가격에 영향을 미치는 변수들을 독립변수로 하여 종속변수인 주택의 가격 등을 예측하는 방법이다.

3.2.3. HEDONIC PRICE FUNCTION 선행연구

HEDONIC PRICE FUNCTION을 통해 주택시장을 분석하는 선행연구로는 교육서비스의 질이 주택가격에 미치는 영향력을 분석한 연구(이주형, 1989), 조망의 영향력을 분석한 연구(정홍주, 1995), 다양한 특성변수의 영향력 비교 및 단기간 가격을 비교한 연구(구본창, 2000) 등 다양한 분야에서 연구가 이루어지고 있다.

3.2.4 HEDONIC PRICE FUNCTION 추정

헤도닉 함수의 여러 가지 함수형태인 선형함수, 이중로그함수, 세미로그함수, 역세미로그함수 4가지 형태로 추정해 본 결과, PRICE보다는 가격의 변화율을 나타내기 위해 종속변수에 로그를 취한 LNPRICE가 종속변수로서 설명력을 나타내는 R2 값이 높기 때문에 세미로그함수를 선택하여 분석을 실시하였다.

선정한 헤도닉 함수의 기본형태는

$$LNPRICE = f ( CPI, M2, GDP, OIL, EXCHANGE, COST, LAND, AREA, PY)이다.$$

AREA는 지역변수로서 DUMMY이고 강남구, 서초구, 송파구, 강동구를 1로 그 외 지역은 0으로 설정하였다. 그 이유는 아파트 분양가격이 지역에 따르는 가중치를 반영해 준 것이며, 위의 4개 구를 선택한 이유는 이미 선행적인 연구의 결과를 토대로 반영하였다. PY변수는 평형이 커지면 커질수록 평당 분양가격이 높아지는 효과를 반영하여 변수로 채택한 것이다. 회귀모형 변수를 입력하여 다중공선성이 발생한 변수인 M2, EXCHANGE, OIL변수를 제거한 후 다중회귀분석을 실행한 결과 다음과 같은 결과가 나타났다.

표 2. 다중회귀분석결과

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.968	.936	.936	.9479

모형	비표준화 계수		표준화 계수		t	유의확률	공선성 통계량	
	B	표준오차	베타				공차한계	VIF
(상수)	-9,377	1,068	-8,782	.000				
PY	8.610E-03	.001	.137	5,909	.000	.922	1,085	
CPI	-.385	.009	.010	-42,001	.000	.163	6,141	
GDP	3.559E-05	.000	.631	17,630	.000	.422	2,369	
COST	4,145E-02	.008	.023	5,454	.000	.333	3,006	
LAND	.502	.006	.031	77,562	.000	.373	2,679	
DAREA	.201	.045	-.512	4,504	.000	.961	1,040	

따라서 추정된 회귀식은 다음과 같다.

$$\begin{aligned} \text{LNPRICE} = & -0.9377 + 8.610\text{E-}03\text{PY} - 0.385\text{CPI} \\ & + 3.559\text{E-}05\text{GDP} + 4.145\text{E-}02\text{COST} \\ & + 0.502\text{LAND} + 0.201\text{DAREA} \end{aligned}$$

분석결과 모형식의 적합도(조정된 결정계수)는 약 94%로 높게 나타났으며, 전체적인 추정계수 중 CPI를 제외한 PY, GDP, COST, LAND 변수는 정(+)의 효과를 가지는 것으로 나타났다. 또한, 지역차이에 대한 터미변수인 DAREA계수가 정(+)으로 나타났으므로 강남구와 서초구, 송파구에 속하는 경우 분양가 상승이 나타나고 있음을 알 수 있다.

### 3.3 신경망분석 결과

#### 3.3.1 분석개요

앞서 분석한 회귀분석의 결과와 비교하기 위하여 신경망분석을 통한 모형을 산출하였다. 신경망 분석을 위한 데이터는 회귀분석에서 도출한 모형에서 독립변수를 input데이터로 하고 종속변수인 LNPRICE를 output데이터로 하여 분석하였다. 본 연구의 목적이 헤도닉모형과 신경망모형의 예측력 비교에 있으므로 동일한 독립변수 사용 시의 예측력을 비교하기 위하여 헤도닉모형에서 다중공선성으로 제외된 변수는 신경망모형의 입력 변수에서도 제외하였다.

신경망분석은 clementine 6.0을 사용하여 실행하였다.(그림 5) 모형생성 시 사용한 설정사항은 신경망생성방법 중 빠른 모드로 하고, 1998년부터 2004년까지의 분양데이터 3,644중 1998년부터 2001년까지의 분양데이터인 2,634개를 훈련자료로 선정하고, 2002년 이후 데이터인 1,010개 데이터를 검정자료로 설정한 후 훈련자료로 신경망 노드를 생성하고 생성된 모형을 검정자료에 적용하였다. 그리고 생성된 예측변수 값을 확인하기 위하여 테이블노드를 연결하여 훈련자료 테이블과 검정자료 테이블을 생성하였다.

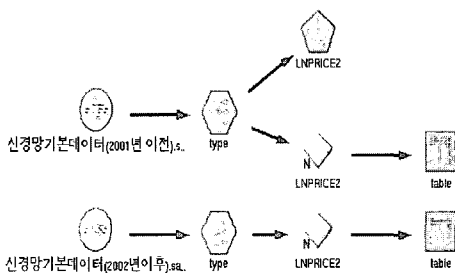


그림 5. 신경망 모형 스트림

#### 3.3.2 분석결과

신경망모형 분석결과 모형의 예측정확도는 97.79%로 상당히 높은 것으로 나타났다. 구체적인 분석결과를 살펴보면 6개의 뉴런이 입력층에 사용되었고, 은닉층에 6개의 뉴런으로 구성되어 있으며, 출력층은 Output 변수가 연속형변수로 구성되어 있으므로 1개의 뉴런으로 구성되어 있다. 모형을 구성하는데 가장 중요한 변수로는 CPI, LAND, PY 순으로 나타났다.

표 3. 신경망분석 결과표

Neural Network "LNPRICE" architecture

Input Layer	: 6 neurons
Hidden Layer #1	: 6 neurons
Output Layer	: 1 neurons
Predicted Accuracy	: 97.79%
Relative Importance of inputs	
CPI	: 0.8193
LAND	: 0.1884
COST	: 0.11727
PV	: 0.10896
GDP	: 0.03628
DAREA	: 0.01563

### 3.4 회귀분석과 신경망분석 결과의 비교

#### 3.4.1 개요

앞서 분석한 회귀분석과 신경망분석의 모형을 통하여 검정데이터를 통하여 예측정확도를 비교 분석한다. 예측의 정확도를 비교하기 위해서 최우선적으로 고려해야 할 점은 추정값이 얼마나 관측값에 근사하는가 하는 문제이다.(이현석 2003) 추정가치 예측의 정확성을 판별하는 기준으로 MAD(mean absolute deviation), RMES(root mean square error), MAPE(mean absolute percentage error), 관측값과 추정값과의 상관분석(correlation), 그리고 관측값에 대한 추정평균값의 단순회귀분석 등이 사용된다.(Kennedy, 1993) 양 모형의 비교 방법으로는 평균제곱오차인 MSE(mean square error)를 사용하였다.1) 이는 신경망 모형이 비선형모형이므로 결정계수가 판정기준으로 부적합하기 때문에 예측값과 실제값의 MSE를 통하여 모형효율성의 판정기준으로 사용하였다.

#### 3.4.2 비교분석결과

2002년1월부터 2004년 1월까지의 분양데이터인 1,010개 데이터의 독립변수, 입력변수를 통하여 결과값을 예측한 결과를 실제값과 비교한 결과 양 모형의 MSE는 다음과 같이 나타났다.

표 4. 비교분석결과

구분	회귀분석	신경망분석
MSE	7.34	0.77

MSE를 통한 예측정확성의 비교분석결과 회귀분석이 MSE가 7.34인데 비하여 신경망분석의 MSE는 0.77로 나타났다. MSE는 추정값과 실제값의 오차를 기준으로 함으로 낮은 값을 가질 경우 예측정확성이 높다고 할 수 있다. 따라서 거시경제 변수를 통해 분양가를 예측하기 위한 모형의 예측정확도를 검증하기 위해 입력데이터를 통한 결과값에 대한 예측에서 신경망모형을 통한 결과의 예측정확성이 우수한 것으로 나타났다.

#### 4. 결론 및 향후과제

본 연구에서는 부동산시장의 분석 및 예측을 위한 방법으로 신경망모형을 소개하고 적용사례를 살펴 보았다. 아직까지 기존 방법을 완전히 대체할 수는 없다. 그러나 몇몇 연구에서 신경망 분석을 통한 예측모형의 정확성도 보였다. 최근 급변하는 시장 환경에서 부동산의 가격, 분양률 등에 대한 예측이 무엇보다 중요하게 되었다. 기존에 제시된 예측모형에 대한 연구는 주로 회귀분석을 이용하였으나, 최근 예측정확도 측면에서 신경망 분석의 우수성이 제시되면서 신경망 분석에 대한 관심이 증가하고 있다. 물론, 신경망 분석을 통한 분석이 기존의 연구 기틀인 회귀분석을 통한 경우보다 이론적 근거가 미약하고 모형도출 과정을 설명하지 못하는 등의 단점을 가지고 있다. 그러나 두 모형을 상호 보완적으로 활용한다면, 부동산분야에 대한 연구는 한 층 진보할 것으로 생각한다. 본 연구의 가장 큰 기여는 기존에 해당 연구를 위해서 단편적으로 소개되었던 신경망에 대한 이론적 배경을 비교적 상세하게 제시한데 있다. 또한, 거시경제변수와 아파트분양가에 대한 관계를 제시하기 위한 모형을 회귀분석과 신경망분석을 통해 산출하고 신경망 분석의 예측정확도의 우수성을 확인하여 그 적용가능성을 제시하였다는데도 의미가 있다.

그러나 부동산 분야에서 신경망분석이 좀 더 활발하게 적용하기 위해서는 몇 가지 향후과제를 해결해야 한다. 먼저, 신경망분석이 더욱 다양한 분야에 대한 연구에서 활용되어야 한다. 신경망 분석의 가장 큰 장점은 모형도출을 위한 제약조건이 많지 않으므로 다양한 분야에서 사용할 수 있다는 점이다. 따라서 좀더 다양한 분야에서 연구가 이루어지고 이에 대한 추가적인 연구가 지속됨으로써, 점차적으로 신경망모형을 적용한 연구의 신뢰성이 높아지고 안정적인 방법론으로 자리를 잡게 될 것이다. 다음으로 해결해야 할 문제는 다른 분석모형과의 연계를 통한 활용

도 제고이다. 특히, 신경망분석은 최근 관심이 고조되고 있는 데이터마이닝(data mining) 분야 중 하나의 분석방법이므로 다른 분석방법인 의사결정나무 분석(decision tree analysis), 군집분석 등 다양한 분석들과 연계하여 분석모형으로서의 활용도를 제고해야 할 것이다.

#### 참고 문헌

- 김태훈 외(2004), 회귀모형과 신경망모형을 이용한 아파트 가격모형에 관한 연구, 국토연구, 제43권, 국토연구원, pp.184~200.
- 전성무(2001), 기업부실예측모형간 성과차이 분석, 홍익대학교 석사학위 논문.
- 정화미 외(2001), 신경망을 이용한 개별공시지가 산정에 관한 연구, 국토연구, 제36권 제7호, 국토연구원, pp55~66.
- 최병남(1994), 인공신경망에 의한 토지평가 시스템, 국토연구, 제22권, 국토연구원, pp37~49.
- 송수미(2001), 데이터 마이닝을 이용한 증권사 위탁고객 이탈 모형구축, 이화여자대학교 석사학위논문.
- 정유석(2003), 인공신경망을 이용한 기업도산예측, 경희대학교 박사학위논문.
- 허준 외(2002), 클레멘타인을 이용한 데이터 마이닝 입문편, SPSS.
- 김대수(1998), 신경망 이론과 응용, 하이테크정보사.
- 이건창(2000), PC를 활용한 회계정보시스템, 무역경영사.
- 이현석(2003), 호텔부동산 가치분석과 사후적 예측정확도 제고방안 연구, 국토계획, 제38권 제6호, 국토연구원, pp23~34.
- 김경태 외(2001), 신경망에 의한 주가수익률 예측모형, 기업경영연구 제14집, 경영연구소 pp127~142.
- 정재호 외(2002), 인공신경망을 이용한 신용카드 부정사용 색출기법에 관한 연구, POSRI 경영연구 제2권 제2호, 경영연구소, pp201~224.
- 김영태 외(2002), 타인을 이용한 데이터 마이닝 입문편, SPSS.
- 이주형(1989), 주거선택 및 주거이동 결정에 학군이 미치는 영향에 대한 연구, 국토연구, 제54권, 국토연구원, pp157~174.
- 정홍주(1995), 아파트가격결정모형에 대한 실증연구, 건국대학교 석사학위논문.
- 구본창(2000), 분양가 차등화를 위한 아파트 특성별 가격차에 관한 연구, 주택연구, 제8권 제2호, 한국주택학회, pp101~123.
- 서승환(1994), 한국 부동산시장의 거시계량분석, 홍문사.

논문제출서: 2005.12.28

심사완료일: 2006.04.12



**〈부 록 : 회귀분석자료〉**

**1. 변수 투입 요소**

진입/제거된 변수

모형	진입된 변수	제거된변수	방법
1	DAREA, GDP, PY, LAND, OIL, COST, CPI	.	입력

- a. 요청된 모든 변수가 입력되었습니다.
- b. 종속변수 : LNPRICE

**2. 모형 요약**

모형	R	R 제곱	수정된 R 제곱	추정값의표준오차	Durbin-Watson
1	.968	.936	.936	.9479	.077

- a. 예측값 : (상수), DAREA, GDP, PY, LAND, OIL, COST, CPI
- b. 종속변수 : LNPRICE

**3. 분산분석**

모형		제곱합	자유도	평균제곱	F	유의확률
1	선형회귀 분석	34238.208	6	5706.368	6350.424	.000
	잔차	2332.715	2596	.899		
	합계	36570.924	2602			

- a. 예측값 : (상수), DAREA, GDP, PY, LAND, OIL, COST, CPI
- b. 종속변수 : LNPRICE

**4. 다중회귀분석 계수**

모형	비표준화 계수	표준 오차	표준화 계수	베타	t	유의 확률	공선성 통계량	VIF
1 (상수)	-9.377	1.068			-8.782	.000		
	PY	8.610E-03	.001	.031	5.909	.000	.922	1.085
	CPI	-.385	.009	-.516	-42.001	.000	.163	6.141
	GDP	3.559E-05	.000	.135	17.630	.000	.422	2.369
	COST	4.145E-02	.008	.047	5.454	.000	.333	3.006
	LAND	.502	.006	.6297	7.562	.000	.373	2.679
	DAREA	.201	.045	.023	4.504	.000	.961	1.040

- a. 종속변수 : LNPRICE

**5. 공선성진단**

모형	차원	고유값	상태지수	분산비율						
				(상수)	PY	CPI	GDP	COST	LAND	DAREA
1	1	6.161	1.000	.00	.00	.00	.00	.00	.00	.01
	2	.724	2.916	.00	.00	.00	.00	.00	.00	.96
	3	.103	7.729	.00	.94	.00	.00	.00	.00	.03
	4	8.676E-03	26.648	.00	.00	.00	.23	.00	.04	.00
	5	2.439E-03	50.253	.00	.01	.06	.50	.02	.10	.00
	6	4.747E-04	113.917	.14	.02	.08	.10	.84	.00	.00
	7	1.661E-04	192.597	.86	.01	.86	.16	.14	.85	.00

**6. 상관계수**

모형		DAREA	GDP	PY	LAND	COST	CPI	
1	상관 계수	DAREA	1.000	.078	-.181	-.032	-.030	-.043
		GDP	0781.000	-.086	-.270	-.235	-.428	
		PY	-.181	-.086	1.000	.152	.065	.032
		LAND	-.032	-.270	.152	1.000	-.331	.749
		COST	-.030	-.235	.065	-.331	1.000	-.582
공분산	DAREA	1.991E-03	7.051E-09	-1.178E-05	-9.377E-06	-1.014E-05	-1.767E-05	
	GDP	7.051E-09	4.074E-12	-2.533E-10	-3.529E-09	-3.608E-09	-7.929E-09	
	PY	-1.178E-05	-2.533E-10	2.124E-06	1.436E-06	7.191E-07	4.236E-07	
	LAND	-9.377E-06	-3.529E-09	1.436E-06	4.186E-05	-1.626E-05	4.445E-05	
	COST	-1.014E-05	-3.608E-09	7.191E-07	-1.626E-05	5.777E-05	-4.059E-05	
CPI	-1.767E-05	-7.929E-09	4.236E-07	4.445E-05	-4.059E-05	8.424E-05		

- a. 종속변수 : LNPRICE

**7. 잔차통계량**

	최소값	최대값	평균	표준편차	N
예측값	4.9584	15.6432	8.4698	3.6275	2603
잔차 표준화	-5.1909	3.9393	8.398E-13	.9468	2603
예측값 표준화	-.968	1.978	.000	1.000	2603
잔차	-5.476	4.156	.000	.999	2603

- a. 종속변수 : LNPRICE

**Abstract**

Neural network analysis is expected to enhance the forecasting ability for the real estate market. This paper reviews definition, structure, strengths and weaknesses of neural network analysis, and verifies the applicability of neural network analysis for the real estate market. Neural network analysis is compared with regression analysis using the same sample data. The analyses model the macroeconomic parameters that influence the sales price of apartments. The results show that neural network analysis provides better forecasting accuracy than regression analysis does, what confirms the applicability of neural network analysis for the real estate market.

**Keywords :** real estate market, neural network, regression model, hedonic price model