

연금복권 판매액 예측모형 개발

최승일

공주대학교 산업시스템공학과

e-mail:ecobath@naver.com

Development of Welfare Lottery Sales Prediction Models

Seungil Choi

Dept of Industrial & Systems Engineering, Kongju National University

요약

본 논문에서는 연금복권 판매액 예측을 위해 기존 연구에서 많이 다루어진 다중회귀모형과 최근 시계열 예측에 활용되고 있는 순환 신경망 모델인 LSTM모형을 활용하여 예측모형을 개발하고, 예측모형에 따른 특징을 살펴보았다. 다중회귀모형은 다양한 설명변수를 추가하여 복권 판매액에 영향을 주는 다양한 요인들을 찾을 수 있으나, 예측을 위해서는 모형의 설명변수에 대한 가정을 최소화하기 위해 국가 경제규모를 나타내는 명목GDP만을 설명변수로 추가하여 모형을 개발하였다. LSTM모형은 미래 전망에 대한 가정을 필요로 하는 설명변수를 추가하지 않고 시계열 자료만을 활용하는 모형을 구성하고, 설명변수가 추가된 다중회귀모형과의 예측력을 비교해 보았다.

2.1 다중회귀모형

1. 서론

연금복권520은 2011년 7월 추첨식 인쇄복권의 판매부진을 개선하고 고액 당첨금으로 인한 사행성을 완화하기 위해 도입되었다. 연금복권720+는 2020년 5월 7일 첫 추첨이 진행되었으며, 연금복권520의 판매부진을 개선하기 위해 연금복권520 출시 이후 경제수준, 물가상승률 등을 감안하여 연금식 복권의 기능을 강화하는 방향으로 당첨구조를 변경하였다.

연금복권 판매액 예측은 제3기 복권사업자 선정을 통해 통합위탁이 이루어진 이후 월별 판매액 (2014년 1월 ~) 자료를 가지고 분석한다. 연금복권720+이 2020년 5월부터 추첨이 이루어져 자료가 충분하지 않으므로, 연금복권520과 연금복권720+의 판매액을 별도로 분석하지 않고 더미변수를 도입하여 하나의 모형으로 분석하고자 한다.

본 연구는 연금복권 판매액의 예측모형을 개발하고 이 모형의 예측력을 검증하여, 연금복권 판매액 예측에 기반한 의사결정에 활용하고자 한다.

2. 연금복권 판매액 예측모형

2.1.1 다중회귀모형 개발

다중회귀분석에 사용한 연금복권 판매액 자료는 33개 분기 (2014년 1사분기 ~ 2022년 1사분기)의 자료이며, 복권 판매액에 영향을 미치는 다양한 요인 중에서 명목GDP를 선택한다 [1-2]. 연금복권 판매액 분기별 계절변동은 명확하지 않으나 명목GDP는 분기별 계절변동을 반영하여 조정한다. 연금복권 판매액은 변동성을 줄여 중장기 예측의 정확도를 높이기 위해 주별 자료를 분기 자료로 변환하고, 분기별 명목GDP는 국가통계포털에서 자료를 확보하여 계절변동을 제거한 후 사용한다. 연금복권720+의 판매시점이 판매점 2020년 4월 14일, 인터넷 2020년 4월 30일을 감안하여 더미변수는 2020년 1사분기까지 0, 2020년 2사분기부터 1로 설정한다.

연금복권 판매액(변수명 WL), 직전 연금복권 판매액(변수명 WL(-1)), 명목GDP(변수명 GDP)는 로그변환하여 다중회귀분석을 실시한다.

$$\ln(WL) = \alpha + \beta_1 \ln(WL(-1)) + \beta_2 \ln(GDP) + \beta_3 DUMMY$$

위 다중회귀식의 계수를 추정하고, 이 식을 이용하여 향후 연금복권 판매액을 추정한다.

2.1.2 다중회귀모형 분석 및 예측

다중회귀모형 분석 결과를 보면 설명변수인 WL(-1),

GDP, DUMMY 사이에 다중공선성이 문제가 되지 않으며, 3개의 설명변수와 상수에 대한 계수 추정 결과가 5% 수준에서 통계적으로 유의함을 확인할 수 있다.

다중회귀모형의 계수 값을 이용하여 27개 분기(2022년 2사분기 ~ 2028년 4사분기)의 연금복권 판매액을 추정한다. 추정에 사용한 설명변수 GDP는 물가상승률을 반영한 명목GDP로 연간 증가율을 4.0%(예측 구간의 물가상승률 전망치를 2%, 실질GDP 성장률 전망치를 2%)로 가정한다. 2014년 1사분기 ~ 2022년 1사분기의 판매액 예측치를 다중회귀모형으로 산출한 후, MAPE(Mean Absolute Percentage Error)를 산출하면 3.11%로 나타난다. 2022년 연금복권 판매액은 1사분기 실측치와 합쳐 226,444,086,000원으로 추정되며, 2023년 ~ 2028년 연금복권 판매액 예측치는 [표 1]과 같다.

[표 1] 연금복권 판매액 다중회귀모형 예측치

	판매액 예측치(원)
2023년	193,593,903,227
2024년	191,175,290,794
2025년	188,922,958,298
2026년	186,697,444,938
2027년	184,498,148,726
2028년	182,324,760,228

2.2 LSTM모형

2.2.1 LSTM모형 개발

파이썬의 케라스(keras)를 이용하여 LSTM(long short-term memory) 기법을 적용하였고, 데이터셋을 다음과 같이 생성한다. 분석은 제3기 복권사업자에 통합위탁이 이루어진 이후 월별 판매액 자료(2014년 1월 ~ 2022년 3월)인 총 99개의 판매액 자료를 사용한다. 연금복권720+이 2020년 5월부터 추첨이 이루어져 자료가 충분하지 않으므로, 연금복권 520과 연금복권720+의 판매액을 별도로 분석하지 않고 더미변수 속성을 추가한다.

판매액 데이터 전처리를 위해 머신러닝에서 많이 사용하는 사이킷런(scikit learn) 라이브러리의 MinMaxScaler를 이용하여 데이터 값을 0과 1사이의 범위로 변환한다. 데이터셋 생성을 위해 종속변수는 월별 판매액으로 설명변수는 직전 12개월의 월별 판매액과 더미변수의 2개의 속성으로 설정하여, 총 87개의 데이터셋을 생성한 후, 사이킷런 라이브러리의 train-test-split를 이용하여 훈련셋과 시험셋을 8:2의 비율로 랜덤으로 구성한다. 케라스를 이용하여 시퀀스 모델을 생성하고, LSTM Layer와 Dropout Layer, Dense Layer를 순차적으로 추가한다. 모델의 학습과정을 설정하기 위해 손실함수와 최적화 방법을 정의하고, 훈련셋을 사용하여 학습을 진행한다.

2.2.2 LSTM모형 분석 및 예측

학습된 모델의 가중치를 이용하여 예측을 하기 위해 predict() 함수를 이용하는데, 초기에 생성한 87개의 데이터셋에서 설명변수를 입력 값으로 한 예측 자료와 실제 자료를 시각화하여 비교하면, 변동성을 줄이며 실제 자료를 추적하는 모습이 나타난다.

2022년 4월 이후 총 81개월의 월별 판매액을 예측하기 위해 실제 자료에 예측 자료를 추가하면서 직전 12개월의 판매액과 더미변수로 구성된 데이터셋으로 다음 주치의 판매액 변환값을 predict() 함수로 추정한다. 함수로 추정한 자료는 0과 1사이의 범위로 나타나는데, MinMaxScaler의 역변환을 위해 inverse_transform() 함수를 이용하여 월별 판매액으로 변환한다. 2022년 1월 ~ 3월 실측치와 2022년 4월 ~ 12월의 예측치를 합하면 181,439,617,000원으로 2022년 연금복권 판매액은 245,411,827,000원으로 추정되며, 2023년 ~ 2028년 연금복권 판매액 예측치는 [표 2]와 같다.

[표 2] 연금복권 판매액 LSTM모형 예측치

	판매액 예측치(원)
2023년	237,572,917,000
2024년	237,141,501,000
2025년	237,121,724,000
2026년	237,120,838,000
2027년	237,120,715,000
2028년	237,120,698,000

3. 결론

연금복권 판매액 예측모형으로 개발한 다중회귀모형과 LSTM모형의 장·단점은 다음과 같다. 다중회귀모형의 장점은 직전 복권 판매액 증가율과 명목GDP 증가율이 복권 판매액 증가율에 미치는 영향을 설명 가능하다는 것이고, 단점은 예측을 위해 설명변수인 명목GDP에 대한 가정이 필요하다는 것이다. LSTM모형의 장점은 예측을 위해 추가되는 설명변수에 대한 가정이 필요하지 않다는 것이고, 단점은 모형을 수립하는 과정에서 무작위 요소와 인자 설정에 따라 모형의 추정계수가 달라질 수 있다는 것과 모형의 계수가 어떤 의미를 갖는지를 설명하기 어렵다는 것이다.

참고문헌

- [1] 이연호, 안병주, 김권, “국제 비교를 통한 복권 판매액 결정요인 분석: OECD 국가를 중심으로”, 경상논집, 제41권 1호, pp. 25-47, 2023년.
- [2] M. J. Kaizeler, H. C. Faustino & R. Marques, "The Determinants of Lottery Sales in Portugal", Journal of Gambling Studies, pp. 729-736, 2014.