



## 장려상

조류인플루엔자의 영향을 고려한 계란가격 예측에 관한 연구

강윤정 · 강소라 · 강수람 · 이하은

# 조류인플루엔자의 영향을 고려한 계란가격 예측에 관한 연구 온라인 매체의 활용을 중심으로

강윤정\*   강소라\*\*   강수람\*   이하은\*\*\*

## Key words

계란가격(price of egg), 조류인플루엔자(AI), 외생변수자기회귀누적이동평균모형(ARIMAX), 벡터오차수정모형(VECM), 장단기메모리모형(LSTM)

## ABSTRACT

Since it first broke out in 1996 in Korea, avian influenza continues to cause domestic egg market distortion. Especially supply-demand imbalance and massive killing due to quarantine measures are leading to social-economic problems. For this reason, it is needed to study the way of predicting egg price which can reflect effects of avian influenza immediately. This study suggests the model for predicting egg price using unstructured data to reflect latest information. Retail price of eggs between January, 2012 and April, 2021 was analyzed with media articles searched for bird flu or avian influenza as an exogenous variable. In addition, avian influenza quarantine history and fresh egg import and export history were used together to enhance the learning power of the model. For prediction, we used statistics models and deep learning model such as ARIMAX, VECM and LSTM. Also, predictive accuracy for each model is presented and compared. LSTM showed best performance between models which recorded lowest MAPE and RMSE. Furthermore the model was able to predict rapid price rising caused by avian influenza.

## 차례

- |                  |            |
|------------------|------------|
| 1. 서론            | 4. 분석 결과   |
| 2. 관련 연구         | 5. 요약 및 결론 |
| 3. 자료의 전처리와 분석방법 |            |

\* 전남대학교 수학/통계학과 석사과정

\*\* 전남대학교 수학/통계학과 석사

\*\*\* 전남대학교 통계학과 학사과정

## 1. 서론

2019년 기준으로 우리나라 농업 생산액의 약 40%를 축산업이 차지하고 있으며, 그 중 계란은 필수적인 식자재로 이용되고 있다. 전염성이 강한 조류인플루엔자의 발생은 농축산업에 부정적인 영향을 미치고 있다. 조류인플루엔자(Avian Influenza, AI)는 닭, 칠면조, 오리 등 조류에서 주로 발생하는 급성 전염병으로 바이러스의 병원성 정도에 따라 저병원성과 고병원성으로 분류된다. 고병원성 조류인플루엔자(Highly Pathogenic Avian Influenza, HPAI)는 위험성이 높은 질병으로 세계동물보건기구에서 이에 대한 발생 사건을 관리하고 있다. 1996년 국내에서 저병원성 조류인플루엔자가 처음 발생하였고 주로 종계의 20%를 폐사하거나 산란율이 감소하는 피해를 야기했다(모인필, 2007). 2003년 HPAI 바이러스는 아시아 전역에 확산되어 국내 최초로 유입되었고, 일부에서는 인체에 감염되어 사망한 사례가 보고되었다. 2006년 HPAI는 육용종계장에서 시작하여 여러 농장에 잇따라 발생하였으나 전파 경로에 대한 정확한 분석이 이루어지지 않았으며, 2009년에는 국내에서 처음 인체에 감염되는 사례가 나타났다. 2014년에서 2015년에는 조류인플루엔자의 발생으로 많은 양의 가금류가 매몰 처리되었고 이로 인한 경제적인 비용이 크게 소요되었다. 특히, 2016년 조류인플루엔자는 대부분 산란계 농장에서 발생하였고 모두 살처분하는 것으로 대응하였으며, 이로 인해 계란의 생산이 급격히 감소하였다.

이와 같이 조류인플루엔자는 국내외적으로 상당한 영향을 미치고 있으나 현재에는 계란 시장에 대한 통계적인 자료와 연구가 많지 않다. 따라서 매년 불안정한 수급과 그에 따른 가격의 급격한 변동을 피할 수 없는 실정이며, 조류인플루엔자의 영향을 고려한 가격 예측 연구의 필요성이 더욱 커지고 있다. 또한, 소비위축의 영향, COVID-19, 예방적 차원의 살처분 등 심리적, 사회적, 정책적인 요인들이 가격에 영향을 미치고 있으며, 이로 인해 소비자 가격을 정확히 예측하는 것에 어려움이 따른다. 종합적으로, 전염성이 있는 가축 질병과 가격에 영향을 미칠 수 있는 요인들을 동시에 고려하여 축산물의 가격을 예측하기 위한 연구들이 필요하다. 기존의 연구방법에서 정형적인 데이터만을 분석하여 빠른 속도로 발전하는 사회를 반영하지 못한다는 한계를 가진다. 정형적인 데이터는 모든 정보를 수집하고 수치화하기까지 시간이 필요하며, 최근 정보를 실시간으로 반영할 수 없다. 최근에는 이러한 한계점을 보완하기 위해

텍스트, 음성, 영상과 같이 비정형 데이터를 정형화하여 활용하는 연구가 진행되고 있다. 특히, 온라인 뉴스 기사는 사회, 경제, 문화, 정치 등 종합적인 이슈의 변화를 빠르게 반영할 수 있다는 점에서 의미 있는 정보를 얻을 수 있다.

지금까지 축산물 전체 또는 일부를 비교 분석한 연구는 많았으나 계란가격만을 집중적으로 분석한 연구는 많지 않았다. 본 연구에서는 조류인플루엔자와 관련한 온라인 뉴스 기사들을 수치화하여 이슈를 실시간으로 반영하고자 온라인 기사 건수와 가격에 영향을 미치는 요인들을 설명변수로 사용하였다. 분석모형으로는 시계열 자료 분석에 주로 이용하는 통계적 모형인 ARIMAX, VECM과 딥러닝 모형인 LSTM을 활용하였다. 결과적으로 조류인플루엔자로 인한 계란가격을 예측함으로써 소비자 물가의 흐름을 파악하고 대응책을 마련하는데 유용한 정보를 제공할 것으로 기대된다. 또한, 농축산물 가격 예측에서 온라인 뉴스 기사와 실측 자료의 결합이 주는 새로운 정보의 의미와 딥러닝 기법의 유용성을 고찰할 수 있다.

## 2. 관련 연구

먼저, 조류인플루엔자의 위험성에 대해 고찰하기 위해 가축 질병과 농산물 가격의 관계에 관한 연구들을 살펴보았다. 박문수 등(2012)은 축산물 가격 변동의 주요 요인이 가축 질병으로 인해 발생한 것으로 보고, 쇠고기 및 돼지고기와 같은 대체재와 비교하여 닭고기와 계란 시장 중심으로 실증분석을 하였다. 쇠고기는 소비위축과 더불어 돼지 살처분의 영향으로 인해 쇠고기 가격이 하락하였고, 돼지고기는 살처분의 마릿수가 증가함으로써 공급이 감소하여 돼지고기의 가격이 상승한 것으로 분석하였다. 김정현 등(2014)은 사건 연구 모형을 통해 구제역 발생이 국내 돼지고기 가격에 미치는 영향을 분석하였다. 3건의 구제역 발생 사건에 대한 일별 자료를 사용하여 구제역이 국내 돼지고기 가격을 하락시키고 하락 추세가 계속된다는 것을 확인하였다. 또한, 대규모 살처분과 같은 정책적 조치는 급격한 공급량의 감소로 이어져 가격이 재상승하도록 유도하여 J 곡선 현상이 나타날 수 있다는 것을 시사하였다. 서대교 등(2014)은 벡터오차수정모형을 통해 구제역과 조류독감 등 가축 질병이 쇠고기, 닭고기, 돼지

고기 가격에 미치는 장·단기적인 영향을 파악하였다. 충격반응함수를 통해 축산물 가격을 예측하였고 다른 축산물 가격과 비교하여 닭고기 가격은 상대적으로 가축 질병 발생에 대한 장기적인 영향을 받았다.

다음으로, 농산물의 가격을 예측하기 위해 활용하는 여러 가지 알고리즘에 대한 선행 연구들을 고찰하였다. 송대식(2001)은 ARIMA 모형을 통해 육계의 가격 예측 모형을 추정하고 가까운 미래를 예측하였다. 향후 1년 동안 갑작스러운 외부충격이 존재하지 않는다면 육계 가격이 전년 대비 하락할 것으로 전망하였다. 명광식(2005)은 Box-Jenkins 시계열 분석 방법을 통해 미래 육계 가격을 예측하였다. 과거의 가격 변화가 미래에도 반복된다는 가정을 두고 적용한다는 점에서 예상치 못한 변화가 일어났을 때 정확성이 떨어질 것이라는 한계점을 제시하였으나 단기와 중기에는 문제가 없을 것으로 판단하여 다른 농산물에 대한 적용이 가능할 것으로 제안하였다. 강태훈(2007)은 AR-GARCH 모형을 통해 축산물의 가격을 예측하고 비선형적 변동성 구조를 살펴보았다. 돼지고기의 가격은 80년대 초반 이후에 감소하는 양상을 보였으나 쇠고기, 닭고기, 계란의 가격은 시간이 지날수록 불안정한 것으로 분석되었다. 여러 부류의 평균 가격을 사용하고 월별 데이터를 사용하였다는 연구의 한계를 지적하고 각 부류에 따른 분석과 함께 주별 데이터나 일별 데이터를 사용할 것을 권장하였다. 최병옥 등(2007)은 VAR, ARIMA, GARCH 모형을 적용하여 과채류의 품종에 적합한 단기 가격을 예측하는 모형을 연구하였다. 생산 대체 관계가 중요한 경우는 VAR 모형이 효율적이었으며, 가격변동의 패턴을 반영하는 것이 중요한 경우는 ARMA 모형이 효율적임을 보여주었다.

마지막으로, 딥러닝 알고리즘을 적용하여 선행된 연구와 농축산물과 관련한 연구들을 살펴보았다. Gim 등(2015)은 광고효과 분석기법을 활용하여 구제역 등 가축 질병에 대한 뉴스가 돼지고기의 수요에 미치는 영향을 농업경제학 입장에서 분석하였다. 김양석 등(2018)은 조류인플루엔자와 관련된 온라인 뉴스 기사 내용에서 키워드 변화를 분석하여 달걀 가격변동과의 상관관계를 알아보았다. 텍스트 마이닝을 통해 키워드의 빈도를 추출하여 시계열에 따른 변화를 분석하고 다중 선형회귀모형을 통해 상승과 하락에 영향을 미치는 키워드를 파악하였다. 신성호 등(2018)은 LSTM 네트워크를 적용하여 자연재해가 농산물의 가격에 미치는 영향을 고려한 예측 모델을 생성하였다. 기상 데이터와 가격 인상률, 수확량, 재배 면적 등의 농산물 자료를 이용하여 구축한 예측모형이 높은 설명력을 보여주었다.

### 3. 자료의 전처리와 분석방법

#### 3.1. 자료의 설명과 전처리

본 절에서는 계란의 가격에 영향을 미치는 요인들을 탐색하였다. 고병원성 조류인플루엔자 발생 시 정부는 방역대를 설정하여 이동을 통제하고 발생 농장에서 사육 중인 가금류를 전부 살처분하는 조치를 한다. 이러한 과도한 살처분으로 공급이 줄어들어 가격이 상승하게 된다. 따라서 계란가격을 예측하고자 할 때 조류인플루엔자의 발생을 가장 중요한 요소로 보고, 이에 대한 심각성을 반영하는 자료를 통해 가격의 변동성을 예측할 수 있다. 다음으로, 온라인 뉴스 기사는 실시간으로 이슈를 반영하기 때문에 조류인플루엔자의 심각성이 높아질수록 언론에서 언급하는 횟수가 증가하게 된다. 방역 건수에 대한 문제가 있더라도 언론을 통한 언급만으로 조류인플루엔자의 발생 예측이 가능하다고 볼 수 있다. 마지막으로, 조류인플루엔자가 심각한 수준에 이르면 정부에서는 수입란을 들여오게 되는데 이 과정에서 관세를 낮추거나 폐지하여 일부로나마 수입란을 들여오려 한다. 정부가 시장에 개입함으로써 가격에 영향을 미치기 때문에 계란의 수입·수출 내역도 가격에 영향을 미치는 요인으로서 역할을 한다고 본다.

따라서 모형에 사용한 자료는 계란가격, 조류인플루엔자 확진 건수·기사 건수, 신선란 수입 무게, 신선란 수출 무게 등이며, 목표변수는 계란가격(Price)이고 독립변수는 확진 건수(Count), 기사 건수(Articles), 신선란 수입 무게(Import), 신선란 수출 무게(Export)로 정의하였다. 다음 <표 1>은 데이터의 수집 기간, 기간 단위, 단위, 출처를 보여준다. 계란가격은 서울지역 대형유통소매점의 일별 특란 가격(30개 기준)이고, 신선란 수입·출입 무게는 월별 신선란 수출입 무게를 수집하였다. 확진 건수와 기사 건수는 웹스크래핑을 통하여 수집하였다. 웹스크래핑은 웹상의 정보를 프로그램을 이용하여 수집하거나 추출하는 기법을 나타내고, 우리는 Python 프로그램의 'Selenium'과 'BeautifulSoup' 패키지를 이용하였다. 모든 자료는 수집 단위가 상이하였으며, 다음과 같은 이유로 주별 단위로 데이터를 가공하였다. 월별 단위로 제공되는 데이터의 수가 많지 않았고 월별 기간의 데이터 정보의 양을 가지고 정확한 정보를

제공하기 어려울 것으로 보았다. 실제 시장에서 월별 예측 가격보다 주별 예측 가격이 소비자들에게 더 좋은 정보를 제공할 수 있을 것으로 판단하였다. 계란가격은 주별 평균으로, 신선란 수출입 무게는 각 월의 주 개수로 나누어 주별 평균으로 가공하였다. 확진 건수와 기사 건수는 주별 합계를 계산하였다.

표 1. 수집 데이터 정보

| 변수명      | 수집 기간     | 기간 단위 | 단위    | 출처            | 비고                              |
|----------|-----------|-------|-------|---------------|---------------------------------|
| Price    | 2001~2021 | 일별    | 원/30개 | KAMIS         |                                 |
| Count    | 2003~2021 | 일별    | 건     | KAHIS         |                                 |
| Articles | 2001~2021 | 일별    | 건     | 네이버<br>(연합뉴스) | “조류인플루엔자” 또는<br>“조류독감”을 포함하는 기사 |
| Import   | 2012~2021 | 월별    | kg    | TRASS         |                                 |
| Export   | 2012~2021 | 월별    | kg    | TRASS         |                                 |

주: 1) KAMIS : 한국농수산물유통공사의 농산물 유통정보 시스템

2) KAHIS : 농림축산검역본부의 국가가축방역통합시스템

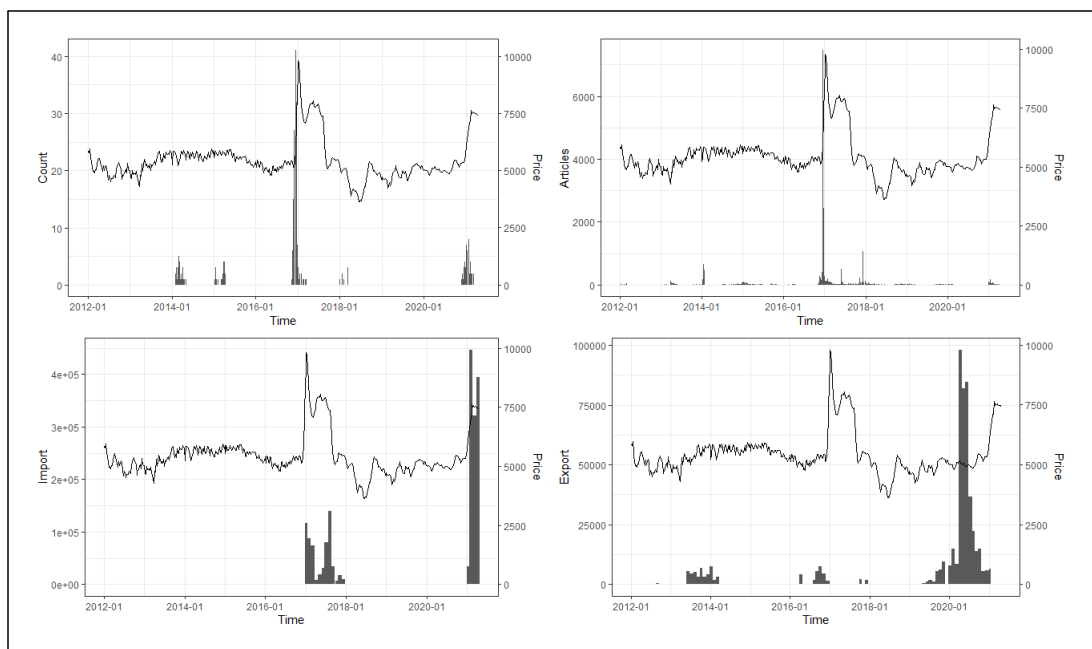
3) TRASS : 한국무역통계진흥원의 한국무역통계정보포털

다음 <표 2>는 계란가격, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 자료의 기초적인 통계 수치를 보여주고 있다. 또한, <그림 1>은 종속변수인 계란가격을 선그래프로, 각 설명변수를 막대그래프로 나타낸 것이다. <그림 1>의 왼쪽 위는 계란가격과 확진 건수를 함께 시각화한 것으로, 계란가격이 오르기 시작할 때 확진 건수가 확연히 많아지는 것을 확인할 수 있다. 이는 <그림 1>의 오른쪽 위에 있는 기사 건수도 같은 모습을 보인다. 특히 조류인플루엔자 확산에 따른 피해가 심각했던 2017년의 경우 계란가격과 확진 건수, 기사 건수가 모두 최고치를 기록하였다. <그림 1>의 왼쪽 아래는 신선란 수입 무게를, <그림 1>의 오른쪽 아래는 신선란 수출 무게를 기록한 그래프이다. 신선란 수입 무게의 경우, 조류인플루엔자가 확산되었던 2017년과 2021년에 가격 안정을 위해 정부가 수입량을 급격히 늘렸다. 신선란 수출 무게의 경우 2020년부터 급격히 늘었다가 최근 조류인플루엔자 확산에 따라 다시 줄어들었다.

표 2. 주별 데이터의 기초 통계

| 구분   | 계란가격   | 확진 건수 | 기사 건수  | 수입 무게    | 수출 무게    |
|------|--------|-------|--------|----------|----------|
| 개수   | 490    | 490   | 490    | 490      | 490      |
| 평균   | 5447   | 0.51  | 44.06  | 15816    | 4261     |
| 표준편차 | 894.90 | 2.88  | 360.87 | 64759.92 | 14267.26 |
| 최솟값  | 3634   | 0     | 0      | 0        | 0        |
| 최댓값  | 9799   | 41    | 7461   | 446393   | 97932    |

그림 1. 계란가격과 설명변수별 시계열 그래프



### 3.2. 분석방법

계란가격을 예측하기 위해 외생변수를 이용한 자기회귀누적이동평균모형(ARIMAX), 벡터 오차수정모형(VECM)과 같은 통계적 시계열모형과 딥러닝모형 중 하나인 장단기 메모리모형(LSTM)을 사용하였다. 일반선형회귀모형은 설명변수의 영향이 시간에 불변한다는 가정을

하고 있어 설명변수의 영향이 변하는 경우 적절하게 반영하지 못한다. 이런 단점을 극복할 수 있는 방법이 ARIMA이다. ARIMA는 현재 관측된 값이 과거의 규칙에 따라 재현된 것으로 보고, 미래에도 이런 규칙이 유지된다고 가정한다. 이 방법은 모형 설정이 편리하지만, 변수 간의 상호작용을 무시하게 된다. 일반선형회귀모형과 ARIMA의 단점과 한계점을 보완하여 제시된 것이 벡터자기회귀모형(VAR)이다. 시계열의 정상성이 만족되지 않으면 VAR을 사용할 수 없지만, 공적분 관계를 가지면 VECM을 적합할 수 있다. 앞서 언급한 시계열 분석 방법의 경우 원 시계열의 길이가 길어질수록 예측의 정확도가 떨어지는 단점이 있다. 딥러닝모형 중 가장 대표적인 시계열 분석 모형인 LSTM의 경우 긴 길이의 데이터에 대해서도 상대적으로 예측 성능이 좋다고 알려져 있다.

### 3.2.1. 시계열 자료의 정상성과 공적분 검정

시계열 자료는 시간이 변화함에 따라 자료에 그 특성이 반영된다. 이 시계열 자료가 불안정할 경우 안정적인 시계열 자료로 전환 후 모형을 추정해야 한다. 불안정한 시계열 자료를 그대로 모형 추정에 사용하면, 아무 상관관계가 없는데도 의미가 있는 것처럼 보이는 허구적 회귀(spurious regression)와 같은 잘못된 결론이 도출되기 때문이다.

시간  $t = 1, 2, \dots, T$ 에 대해서 다변량 시계열이 아래 조건을 만족하는 경우 정상성(stationarity)을 만족한다고 한다.

$$(1) E(X_t) = \mu < \infty$$

$$(2) Var(X_t) = E[(X_t - \mu)(X_t - \mu)'] = \Sigma_X < \infty$$

$$(3) Cov(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)'] = T_X(k)$$

이는 평균과 분산, 공분산이 시간에 따라 불변함을 의미하는데, 정상시계열은 일시적인 충격에 의해 추세치를 이탈하더라도 결국 추세에 회귀하는 경향을 가진다. 시계열의 정상성을

확인하기 위해 단위근 검정을 시행한다. 본 연구에서는 일반적으로 널리 사용하고 있는 ADF (Augmented Dickey-Fuller test) 검정 방법을 사용하였다.

$$(4) \Delta y_t = \alpha + \delta y_{t-1} + \sum_{j=1}^p \gamma_j \Delta y_{t-j} + \epsilon_t$$

위 식에서  $y_t$ 는 검정할 시계열 변수,  $\delta$ 는 검정계수,  $p$ 는 시차,  $\epsilon_t$ 은 백색잡음인 오차항을 나타낸다. ADF 검정의 귀무가설  $H_0$ 은  $\Delta y_t \sim I(0)$ 이다.  $\delta$ 가 0이 되면 귀무가설을 기각하는데, 이는 단위근이 존재하지 않는 안정적인 시계열임을 의미한다. 귀무가설을 기각하지 않는 반대의 경우, 시계열이 불안정함을 의미한다.

개별 시계열이 비정상(불안정)하여 변동이 있더라도 이들 사이에 장기적으로 안정적인 시계열을 생성하는 선형결합이 존재한다면 이 선형결합은 정상이며, 이들 시계열은 공적분(cointegration) 관계에 있다고 한다. 비정상시계열 사이에 공적분이 존재한다면 이를 나타내는 벡터오차수정모형을 이용해 모형을 추정할 수 있다. 공적분 검정을 하는 방법은 앵글&그레인저 공적분 검정(Engle and Granger Cointegration test)과 요한슨 공적분 검정(Johansen test)이 있다. 본 연구에서는 앵글&그레인저 공적분 검정의 한계점을 극복하고 다변량 시차의 구조를 설명하기에 적합한 요한슨 공적분 검정을 수행하였다.

### 3.2.2. 외생변수를 포함한 자기회귀누적이동평균모형(ARIMAX)

회귀모형은 시간의 흐름과 관계없이 설명변수의 영향력이 일정하다는 한계를 가지며, 이를 보완한 방법이 시계열 모형이다. 자기회귀모형(AR)은 임의의 변수에 대해  $t-p$ 시점 관측값이  $t$ 시점 관측값에 영향을 미친다는 가정을 활용하고, 이동평균모형(MA)은  $t-q$ 시점 백색잡음과  $t$ 시점 백색잡음으로 추세를 파악하여 미래를 예측한다. 자기회귀이동평균모형(ARMA)은 AR과 MA이 결합된 형태로 정상성을 만족하는 자료에 적합한 방법이다. ARIMA는 비정상적인 시계열 자료를 분석하는 기법으로 자기 자신의 추세만을 고려하는 일변량 자기회귀모형이다. 따라서 장·단기적인 예측력을 높이기 위해서 시간의 흐름에 따른 여러 가지 외부정보를 반영할 필요가 있다. ARIMAX(Auto Regressive Integrated Moving Average

with Exogenous)는 ARIMA에 독립변수를 추가하여 외생변수 또는 내생변수의 영향을 고려하는 다변량 자기회귀모형이다. ARIMAX(p,d,q,r) 모형의 식은 다음과 같다.

$$(5) \quad y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \beta X$$

여기서  $c$ 는 상수,  $\epsilon_t$ 와  $\epsilon_{t-1}$ 는 동일한 두 기간에 대한 오류항,  $p$ 는 자기회귀모형의 차수,  $d$ 는 적분차수,  $q$ 는 이동평균 모형의 차수,  $r$ 은 독립변수의 지연 차수를 의미한다.

### 3.2.3. 벡터오차수정모형(VECM)

벡터오차수정모형(Vector Error Correction Model, VECM)은 오차수정모형을 VAR 모형으로 확장한 것이다. 공적분 검정을 통해 변수 간 공적분이 존재가 확인되면 장기적인 예측에 앞서 오차수정모형을 구성하여 추정할 필요가 있다. VECM은 특정 시점에서 장기적인 균형의 괴리가 발생하면 균형의 오차가 시간의 흐름에 따라 조정된다는 개념으로부터 설계된 모형이다. 앵글&그레인저(1987)는 두 변수가 비정상적인 시계열이면서 공적분 관계에 있을 때, 식 (6)을 추정하기 위해 식 (7)과 같은 오차수정모형으로 바꿔 쓸 수 있으며, 여기서  $EC_{t-1}$ 은 오차수정항이 된다.

$$(6) \quad y_t = \alpha + \lambda y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

$$(7) \quad (y_t - y_{t-1}) = \alpha - (1 - \lambda)y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1)x_{t-1} + \epsilon_t$$

$$\begin{aligned} \Delta y_t &= \beta_0 \Delta x_t - (1 - \lambda) \left[ y_{t-1} - \frac{\alpha}{1 - \lambda} - \frac{(\beta_0 + \beta_1)}{1 - \lambda} x_{t-1} \right] + \epsilon_t \\ &= \beta_0 \Delta x_t - (1 - \lambda) EC_{t-1} + \epsilon_t \end{aligned}$$

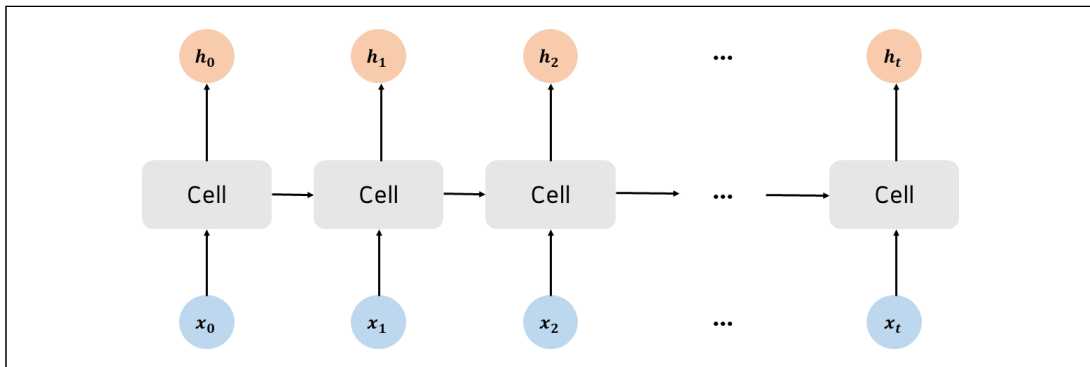
공적분 관계를 가지는  $y_t, x_t$ 의 오차수정항은 정상시계열이다. 식 (7)에 대한 OLS 추정량은 일치추정량이 되고  $t$ 검정이 가능하다.  $y_{t-1}$ 이  $x_{t-1}$ 의 균형값에 비해 크면 균형 오차는 항상

양이 되고 다음 시점인  $y_t$ 에 대해 음의 조정이 작용된다. 오차수정항의 계수인  $(1 - \lambda)$ 는 장기 균형으로 접근하는 조정속도이다. 안정적인 장기 균형을 위해  $(1 - \lambda)$ 는 0보다 커야 한다.

### 3.2.4. 장단기메모리모형(LSTM)

순환신경망모형(Recurrent Neural Network, RNN)은 연속성을 가지면서 시점 간의 연관성이 있는 데이터를 분석하는데 유용하다. 신경망 내부의 재귀적 특성으로 과거 시점의 데이터와 현재 시점의 데이터의 관계를 기억하여 순서 혹은 시간에 따른 데이터의 변화를 분석하는데 유용하다. <그림 2>는 RNN 구조로 새로운 정보를 입력받는 동시에 각 셀에서 다음 셀로 정보가 전달된다.  $x_0, x_1, \dots, x_t$ 는 각 시점에 대한 새로운 입력값이고  $h_0, h_1, \dots, h_t$ 는 그에 대한 출력값이다.

그림 2. RNN 구조



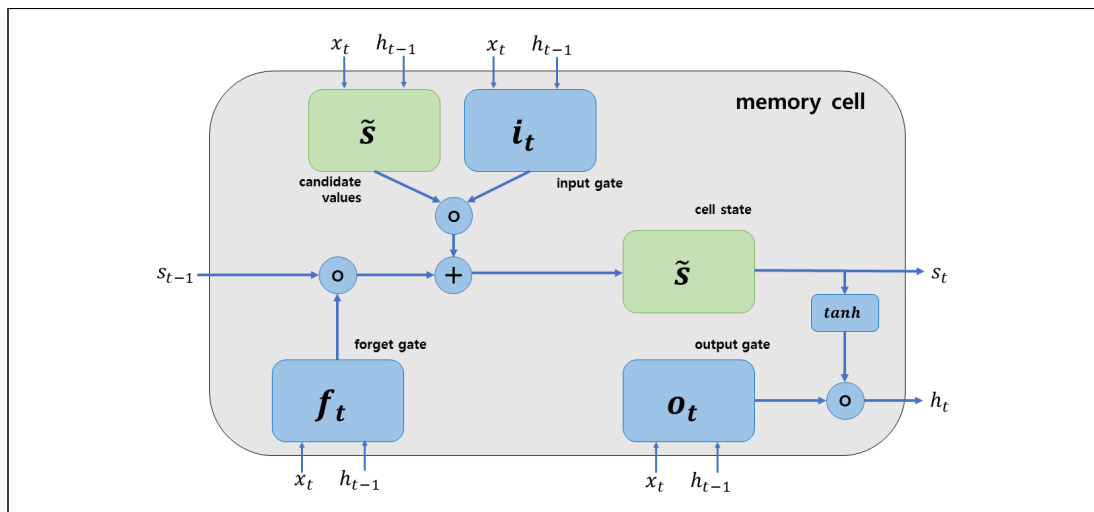
RNN에서 가장 중요한 점은 새로운 입력을 학습하면서 입력된 정보가 출력값의 형태로 다음 셀에 전달되어 신경망 내부에서 순차적 흐름의 학습이 동시에 이루어진다는 점이다. 이러한 구조로 시간에 따른 데이터를 학습하고 분석하는데 유용하게 이용되고 있다.  $t$ 시점의 출력값인  $h_t$ 는 식 (8)과 같이 나타낼 수 있다.

$$(8) \quad h_t = \tanh(h_{t-1} W_h + x_t W_x + b)$$

RNN에서  $t$ 시점의 출력값  $h_t$ 는 두 개의 가중치가 있다.  $Wx$ 는 입력값인  $x_t$ 를  $h_t$ 로 변환시키는 가중치이고  $W_h$ 는 이전 단계의 출력값인  $h_{t-1}$ 을 다음 단계로 전달하는데 변환시키는 가중치로 학습을 통해 업데이트가 이루어진다.  $b$ 는 바이어스를 나타내며 이들은 활성화함수에 의해 변환되어 출력값인  $h_t$ 가 생성된다. 그러나 RNN은 입력 데이터가 길어질수록 과거의 데이터의 영향이 감소하는 기울기 소실(vanishing gradient)이 발생하여 가중치 갱신이 일어나지 않고 학습의 문제가 생기는 장기 의존성 문제가 나타난다. LSTM은 이러한 문제에 대한 대안으로 나타났다.

LSTM은 RNN과 같이 순차적으로 정보가 전달되어 재귀적 특성을 갖지만, 신경망 내부에 메모리 셀이라는 개념이 추가되었다. 메모리 셀은 각 시점의 cell state를 통해 일정 비율로 기억하거나 폐기함으로써 과거의 정보를 얼마나 반영할지 제어할 수 있다. 이를 통해 RNN의 장기 의존성 문제를 해결하여 먼 시점 간의 관계를 학습하거나 그 정도를 조절할 수 있다. 메모리 셀 내부에는 forget gate, input gate, output gate라는 요소가 이전 정보와 입력 정보, 출력정보를 유지 및 조절한다. 여기서 forget gate는 cell state에서 정보의 망각 정도, input gate는 정보의 반영 정도, output gate는 출력을 결정한다. 각 게이트는 sigmoid 연산을 통해 0에서 1 사이의 값을 출력한다.

그림 3. LSTM Cell 구조



아래의 식 (9)의 forget gate에서는 과거의 정보를 얼마나 반영할 것인지를 결정하고 식 (10)의 input gate에서는 현재의 정보를 얼마나 반영할 것인가를 결정한다. 식 (11)은 활성화 함수를 통해 cell state에 반영할 값을 나타내고 식 (12)는 앞의 과정이 반영된 현재 시점의 cell state이다. 식 (13)의 output gate는 출력값을 나타내며 식 (14)는 output과 cell state의 조합으로 다음 셀에 전달할 값이다.

$$(9) f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$(10) i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$(11) \tilde{s}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$(12) s_t = f_t \cdot s_{t-1} + i_t \cdot \tilde{s}_t$$

$$(13) o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$(14) h_t = o_t \cdot \tanh(s_t)$$

### 3.2.5. 성능지표

예측 모델 간의 성능 비교를 위한 기준으로 회귀문제에서 대표적으로 쓰이는 두 가지 지표인 MAPE(Mean Absolute Percentage Error)와 RMSE(Root Mean Square Error)를 사용하였다. MAPE는 잔차에 대한 실제값의 비율을 의미하며 백분율로 나타낸다. 실제값이 1보다 작거나 0인 경우에는 사용하기 어렵지만 본 연구의 데이터는 값이 비교적 크고 0을 갖지 않기 때문에 적용하기에 알맞다. RMSE는 잔차의 제곱값을 평균한 MSE에 제곱근을 취한 것으로 예측변수와 같은 단위로 해석할 수 있다. 잔차를 제곱하기 때문에 이상치에 민감한 편이다. 두 지표의 공식은 다음과 같다.

$$(15) MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

## 4. 분석 결과

본 연구에서는 2012년부터 2018년까지의 데이터는 모형의 학습에 사용하였고, 2019년부터 2021년까지의 데이터는 학습된 모형의 성능을 평가하는데 사용하였다. 설명변수의 단위가 다르고, 값의 크기가 달라 절대적 크기의 차이로 인한 모형의 학습이 왜곡되는 것을 방지하고자 표준화하여 사용하였다. 각 설명변수의 평균과 표준편차를 이용하여 절대적 크기의 차이를 줄였다.

### 4.1. 시계열 자료의 정상성 및 공적분 검정 결과

시계열 자료의 정상성 확인을 위해 수행한 ADF 검정 결과는 아래 <표 3>과 같다. 검정 결과 계란가격과 신선란 수입 무게는 귀무가설을 기각하지 못해 단위근을 가지는 불안정 시계열이었다. 원 데이터를 1차 차분하면 유의수준 1%에서 귀무가설을 기각하여 정상적인 시계열임을 확인하였다.

표 3. ADF 검정 결과

| 구분       | 원데이터              | 1차 차분 데이터        |
|----------|-------------------|------------------|
| Price    | -1.85(0.63)       | -11.18(0.01)***  |
| Count    | -6.7016(0.01)***  | -15.902(0.01)*** |
| Articles | -15.329(0.01)***  | -24.353(0.01)*** |
| Import   | -1.8891(0.6251)   | -15.657(0.01)*** |
| Export   | -3.964(0.0108)*** | -15.541(0.01)*** |

주: \*, \*\*, \*\*\*은 각각 유의수준 10%, 5%, 1%에서 유의함을 의미함.

또한, 앞에서 언급한 바와 같이 개별 시계열 자료들이 비정상적이더라도 변수 간의 선형 결합이 안정적이면 그 변수 간에는 장기 균형 관계가 존재하게 된다. 요한슨 공적분 검정은 시계열 변수들 사이에 존재하는 공적분 벡터의 수를 우도비(likelihood ratio) 검정을 통해 결정하고 이후 VECM을 추정한다. 공적분 검정을 하기에 앞서 적절한 VAR의 차수를 결정해야 한

다. 차수를 결정하는 기준은 AIC(Akaike), SC(Schwarz) 등의 기초하여 결정하는데, AIC는 시차가 크게 선택되는 경향이 있어 SC를 기준으로 적정 시차를 2로 설정하였다. 아래 <표 4>에서 공적분 검정 결과를 확인할 수 있고, 유의수준 5%에서 4개 이상의 공적분 관계가 있으므로 VECM으로 분석이 가능하다.

표 4. 공적분 검정 결과

| $H_0$      | Eigenvalue | Trace Statistic | 5% Critical Value |
|------------|------------|-----------------|-------------------|
| $r = 0$    | 0.4618     | 417.60          | 70.60***          |
| $r \geq 1$ | 0.3187     | 190.85          | 48.28***          |
| $r \geq 2$ | 0.0585     | 50.36           | 31.52***          |
| $r \geq 3$ | 0.0467     | 28.29           | 17.95***          |
| $r \geq 4$ | 0.0289     | 10.75           | 8.18***           |

주: \*, \*\*, \*\*\*은 각각 유의수준 10%, 5%, 1%에서 유의함을 의미함.

## 4.2. ARIMAX 추정 결과

ARIMAX를 적합하기 위해서는 적절한 계수를 선택해야 한다. 이를 위해 R 프로그램의 forecast 패키지를 이용하였다. AIC, BIC 등을 이용해 적절한 계수를 선택할 수 있는데, 계란가격을 추정하기 위한 적절한 모형으로는 ARIMAX(2,1,0)가 선택되었다. 추정된 모형은 아래와 같으며 <표 5>에서 각 계수별 표준편차와 유의확률을 확인할 수 있다. 각 계수를 확인하면 AR(1)과 AR(2)가 유의수준 1%에서 유의하며 시점에서 계란가격에 양의 영향을 주고 있다. 확진 건수 역시 유의한 변수이나 자기회귀항과는 다르게 음의 영향을 가지는 것을 확인할 수 있다. 이외의 기사 건수, 신선란 수입 무게, 신선란 수출 무게는 유의하지 않았다.

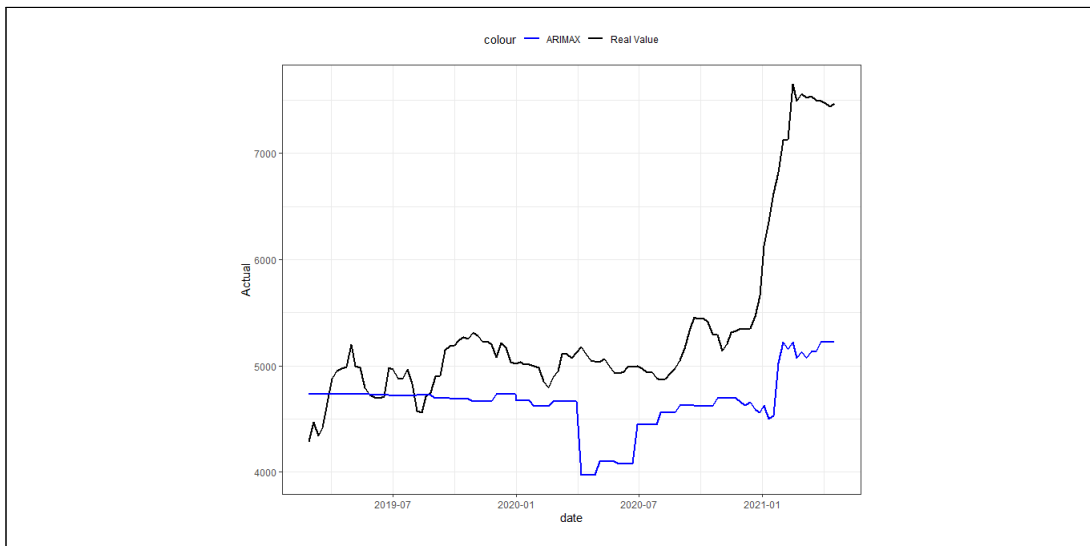
표 5. ARIMAX 모형 추정 결과

|           | Estimate | std.Error | z-value | Pr(> z )     |
|-----------|----------|-----------|---------|--------------|
| $y_{t-1}$ | 0.2194   | 0.0556    | 3.94    | 8.148e-05*** |
| $y_{t-2}$ | 0.1573   | 0.0528    | 2.97    | 0.002896**   |

|          | Estimate  | std.Error | z-value | Pr(> z )     |
|----------|-----------|-----------|---------|--------------|
| Count    | -96.6296  | 15.4355   | -6.26   | 3.845e-10*** |
| Articles | 11.3210   | 6.9668    | 1.625   | 0.1041       |
| Import   | 80.5537   | 66.3087   | 1.2148  | 0.2244       |
| Export   | -110.8116 | 180.3068  | -0.6146 | 0.5388       |

아래 <그림 4>는 추정된 모형을 이용해 예측한 값을 그린 것이다. 추정된 모형인 ARIMAX (2,1,0)는 평균적인 가격을 낮게 예측하였으며 변동을 정확히 포착하지 못했다. 예측된 계란 가격은 2020년 4월 초에 수직으로 하락하고 이후 10주가량 하락한 가격으로 유지되다가 2020년 6월 초에 다시 원래의 가격으로 급격히 상승하는 모습을 보인다. 이는 이전까지 꾸준히 조류인플루엔자 관련 기사가 보도되었지만 2020년 2월부터 급격히 퍼지기 시작한 COVID-19가 사회적으로 이슈화되어 조류인플루엔자에 대한 관심이 줄어들어 기사 건수가 0인 것의 영향으로 볼 수 있다. 이후 2020년 말부터 서서히 조류인플루엔자가 발병하며 2021년에 들어서면서 본격적으로 확산되어 계란가격이 급격히 상승하는 모습을 볼 수 있다. 실제로는 가격이 3,000원 정도 상승하였으나 예측값은 500원 정도에 그치는 것을 확인하였다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.1213, 961.8527이었다.

그림 4. 실제 계란가격과 ARIMAX 예측값 비교



### 4.3. VECM 추정 결과

VECM은 예측하고자 하는 계란가격과 다른 변수를 함께 모형에서 사용하므로 따로 표준화나 정규화를 하지 않은 원 데이터를 사용하였다. VECM 추정 결과는 아래 <표 6>과 오차수정항은 아래 식 (17)과 같다.

표 6. VECM 추정 결과

|                         | $\Delta Price$  | $\Delta Count$   | $\Delta Articles$ | $\Delta Import$  | $\Delta Export$  |
|-------------------------|-----------------|------------------|-------------------|------------------|------------------|
| $ECT_{t-1}$             | -0.002(0.0043)  | 0.000(0.00)      | -0.0022(0.0066)   | 0.3967(0.185)*   | 0.0574(0.016)*** |
| <i>Intercept</i>        | 7.1079(19.29)   | -0.069(0.136)    | 9.283(29.529)     | -1474.47(832.4)  | -222.36(75.67)** |
| $\Delta Price_{t-1}$    | 0.26(0.054)***  | -0.001(0.0004)** | 0.0998(0.082)     | 10.617(2.335)*** | -0.2576(0.2123)  |
| $\Delta Count_{t-1}$    | 10.3573(8.279)  | 0.4207(0.061)*** | 124.45(13.35)***  | -637.022(376.6)  | -70.939(34.23)*  |
| $\Delta Articles_{t-1}$ | 0.0682(0.0369)  | -0.002(0.000)*** | -0.801(0.056)***  | -3.867(1.5945)*  | -0.1756(0.1449)  |
| $\Delta Export_{t-1}$   | 0.0149(0.0134)  | -0.0002(0.00)    | 0.0473(0.0205)*   | 0.3048(0.577)    | 0.0202(0.0525)   |
| $\Delta Import_{t-1}$   | 0.0035(0.001)** | 0.000(0.000)     | -0.0015(0.0019)   | 0.0184(0.0525)   | 0.0051(0.0048)   |
| $\Delta Price_{t-2}$    | 0.1812(0.055)** | -0.0009(0.000)*  | 0.1966(0.0843)*   | 5.0292(2.3772)*  | -0.3543(0.2161)  |
| $\Delta Count_{t-2}$    | 13.3292(10.90)  | 0.0435(0.0769)   | 135.44(16.69)***  | -81.485(470.64)  | 87.313(42.782)*  |
| $\Delta Articles_{t-2}$ | 0.0099(0.040)   | 0.0005(0.0003)   | -0.312(0.061)***  | 6.4895(1.724)*** | -0.379(0.1568)*  |
| $\Delta Export_{t-2}$   | -0.0032(0.013)  | 0.000(0.000)     | 0.0676(0.0206)    | -0.0437(0.5805)  | 0.0269(0.0528)   |
| $\Delta Import_{t-2}$   | -0.0024(0.001)* | 0.000(0.000)     | -0.018(0.0017)    | -0.013(0.0471)   | 0.000(0.0043)    |

$$(17) ECT_{t-1} = Price - 155.86*Count - 0.70*Articles - 0.08*Import - 1.07*Export$$

식 (17)에서 계란가격, 확진 건수, 기사 건수, 신선란 수출입 무게의 장기균형 관계를 확인할 수 있다. 오차수정항에서 오차항이 0이면,

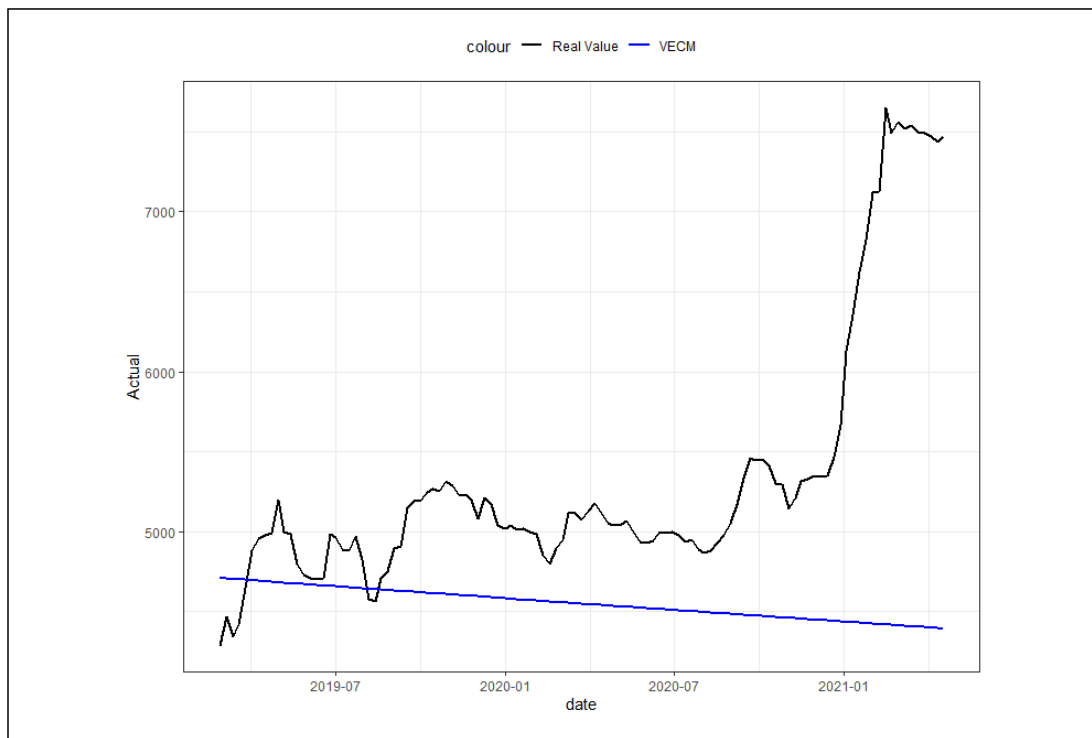
$$(18) Price = 155.86*Count + 0.70*Articles + 0.08*Import + 1.07*Export$$

식 (18)과 같이 정리할 수 있다. 이는 확진 건수가 1% 증가하면 계란가격은 155.86% 증가하고, 기사 건수, 신선란 수입 무게, 신선란 수출 무게가 각각 1%씩 증가하는 경우 계란가격

은 0.7%, 0.08%, 1.07% 증가함을 의미한다. 확진 건수가 가장 영향력이 컸으며 신선란 수입 무게는 다른 변수에 비해 미미한 것으로 나타났다. 오차수정항( $ECT_{t-1}$ )의 계수를 살펴보면 다른 변수에 비해 신선란 수입 무게(0.3967)의 절댓값이 가장 큰 것을 볼 수 있는데, 이는 만약 장기 균형 관계에 문제가 생긴다면 신선란 수입을 통해 계란가격의 조정이 이루어짐을 알 수 있다.

추정된 모형을 이용해 2019년부터 2021년까지의 주별 계란가격을 예상한 결과는 <그림 5>와 같다. 2019년까지는 평균적인 선을 잘 따라가는 것처럼 보이나 2021년으로 갈수록 점점 예측력이 떨어지는 것을 볼 수 있다. 특히 조류인플루엔자가 발생해 계란가격이 상승하기 시작한 2020년 말부터는 치솟는 계란가격을 따라가지 못했다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.1352, 1175.4238이었다.

그림 5. 실제 계란가격과 VECM 예측값 비교



#### 4.4. LSTM 추정 결과

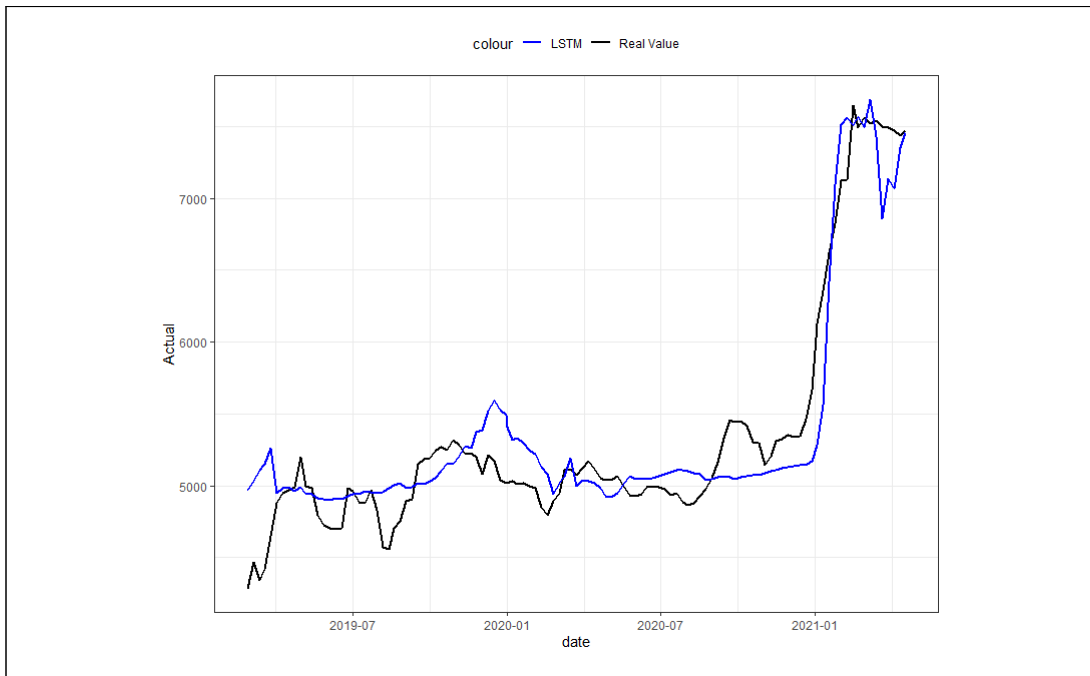
LSTM을 이용하여 시계열 분석을 하기 위해서는 연속된 형태의 데이터 구조를 시퀀스 형태로 변환하는 과정이 필요하다. 분석에 사용되는 데이터는 2012년 1월에서 2021년 4월까지 주 단위로 연속적인 형태이다. 이를 목적과 모델의 성능에 따라 학습 sampling window와 target의 형태로 데이터를 구조화해준다. LSTM은 기간의 데이터를 학습하여 target의 데이터를 예측하는 형태가 된다. 본 연구에서는 sampling window를 6주, 8주, 10주, 12주로 하였고 target을 1주로 하여 최소 6주에서 최대 12주의 데이터를 학습하여 다음 한 주의 가격을 예측하는 네 개의 데이터셋으로 구조화하였다. 또한, 계란가격의 연간 주기성을 반영하기 위하여 매년의 주 번호(weeknumber)를 변수로 추가하여 주 번호, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 등 다섯 가지의 설명변수를 이용하였다.

LSTM의 하이퍼 파라미터를 설정하기 위하여 그리드 탐색을 하였고 예측 성능이 가장 좋은 값으로 최적화하였다. 설정한 하이퍼 파라미터는 은닉층과 은닉노드의 개수, dropout과 recurrent dropout 비율, 최적화 알고리즘, 학습률, 에포크, 배치 크기 등이다. 분석 결과 가장 좋은 모델은 sampling window가 8주이고 각각 32개와 16개의 은닉노드를 갖는 2개의 은닉층을 갖는 모델로 나타났다. 과적합을 방지하는 방법으로써 LSTM에 적용되는 dropout은 각 cell의 input에 대해 수직 방향으로 이루어지는 dropout과 이전 셀에서 현재 셀로 전달되는 hidden state에 대해 수평 방향으로 이루어지는 recurrent dropout이 있다. recurrent dropout의 경우 본 연구 모델에 적용 시 성능을 하락시키므로 적용하지 않았고, dropout의 경우 input의 영향을 다소 감소시켜 성능을 향상시키므로 두 개의 은닉층에서 각각 0.2의 비율로 적용하였다. 최적화 알고리즘은 adam을 사용하였으며 학습률은 0.01이고 에포크는 100번, 배치 크기는 16으로 학습하였다.

〈그림 6〉은 추정된 모형을 이용해 2019년부터 2021년까지의 주별 계란가격을 예상한 결과이다. 예측 초기에는 실제값과 예측값의 방향과 값의 차이가 다소 크게 나타났지만, 이후에는 값과 방향의 차이가 크지 않는 것으로 나타났다. 2021년의 경우 실제 계란가격이 최고점에 다다른 뒤 횡보하고 있지만 예측값은 한번 떨어졌다가 복귀하는 모습을 보이고 있다. 이는 앞서 설명했던 ARIMAX 추정 결과와 같이 기사 건수의 왜곡에서 비롯되는 것으로 추측해볼

수 있다. 마지막으로 추정된 모형을 이용해 계산한 MAPE와 RMSE는 각각 0.0412, 282.7135 이었다.

그림 6. 실제 계란가격과 LSTM 예측값 비교



## 5. 요약 및 결론

계란은 다양한 용도로 활용되고 있는 필수 식자재 중 하나이다. 전염성이 강한 조류인플루엔자의 발생과 가금류 농장에 대한 방역 조치는 안정적인 계란 수급과 생산에 어려움을 불러 일으킨다. 고병원성 조류인플루엔자는 급성 전염병으로 폐사하거나 산란율이 감소하는 등의 피해를 야기한다. 전염성이 높기 때문에 예방적 차원에서 주변 농장의 가금류를 살처분함으로써 대응하는데, 이는 계란가격의 상승으로 이어진다. 따라서 조류인플루엔자의 영향을 고려한 계란가격 예측 연구가 필요하다. 정형적 정보는 정보 수집과 수치화까지 상대적으로 긴 시

간이 필요한데, 온라인 뉴스 기사는 종합적인 이슈와 변화를 빠르게 반영할 수 있다는 장점이 있다. 본 연구에서는 조류인플루엔자와 관련한 뉴스 기사를 수치화하고, 계란가격에 영향을 미칠 수 있는 요인들을 설명변수로 사용하여 계란 주별 소매 가격의 변동을 예측하고자 하였다.

분석에 사용한 데이터는 계란가격, 확진 건수, 기사 건수, 신선란 수입 무게, 신선란 수출 무게 등이며 계란가격을 종속변수로 사용하였다. 월별 단위로 제공되는 데이터 분석의 한계와 실제 소비자에게 정보 효용성을 고려하여 주별 단위로 데이터를 가공하고 분석에 사용하였다.

분석 방법으로는 ARIMAX, VECM 그리고 LSTM을 사용하였다. 분석에 앞서 원데이터의 정상성과 공적분 검정을 수행하였다. 계란가격과 신선란 수입 무게는 불안정 시계열이었으며 이외 확진 건수나 기사 건수, 신선란 수출 무게는 정상시계열이었다. 본 연구의 분석 데이터와 같이 비정상시계열이면서 공적분 관계를 가지면 VECM을 사용할 수 있다. 2012년부터 2018년까지의 데이터를 이용해 모형을 추정하고 2019년부터 2021년까지 데이터를 이용해 추정된 모형의 성능을 평가하였다. 가장 좋은 성능을 보인 것은 LSTM이었으며, 다른 모형에 비해 뛰어난 결과(MAPE 0.042, RMSE 282.7135)를 보였다. 전반적인 추세와 가격의 급격한 변동 모두 포착하였으나 본 연구의 설명변수로 확인할 수 없는 주기적 변동은 예측이 불가능하였다. ARIMAX의 경우, 전체적인 추세는 잘 예측하였지만 조류인플루엔자로 인한 급격한 가격 변화는 예측하지 못하였다. 이는 설명변수의 영향이 잘 반영되지 못한 것으로 보인다. VECM의 경우, 전반적인 추세와 방향을 모두 예측하지 못하므로 계란가격 예측을 위한 모형으로 적합하지 않았다. 통계적 시계열 모형은 장기적 예측과 급격한 변동을 가지는 데이터를 분석하는데 한계를 가진다. LSTM은 상대적으로 좋은 성능을 보였으나 계란가격을 예측하기 위한 변수가 충분하지 않았다. 따라서 이를 보완할 수 있는 다른 변수를 찾는 것이 향후 연구에 중요한 과제가 될 것이다.

## 참고문헌

- 모인필. 2007. “고병원성 조류인플루엔자의 역학적 특성과 국내 발생 동향.” 『Crisisonomy』 3(2): 30-37.
- 박문수, 황선웅, 서상택. 2012. “축산물 시장의 가격전달채널 분석 및 가격안정화 방안: 육계 및 계란 시장을 중심으로.” 『한국농촌경제연구원 연구자료』 1-158.
- 김정현, 서병선. 2014. “구제역 발생이 돈육가격에 미치는 영향: 사건 연구(Event Study)의 적용.” 『농촌경제』 37(1): 97-114.
- 서대교, 신종협. 2014. “가축질병이 축산물가격에 미친 영향 분석.” 『산업경제연구』 27(6): 2431-2450.
- 송대식. 2001. “육계의 가격예측과 가격변화에 관한 연구.” 『식품유통연구』 18(1): 117-130.
- 명광식. 2005. “Box-Jenkins 모형을 이용한 육계가격 예측.” 『농촌경제』 28(2): 73-83.
- 강태훈. 2007. “주요 축산물 가격변동의 시계열적 특성.” 『농업경영·정책연구』 34: 369-388.
- 최병옥, 최익창. 2007. “시계열 분석방법을 이용한 과채류 월별가격 예측.” 『농촌경제』 30(1): 129-148.
- Uhn-Soon Gim, Se-Hyun Choi, and Jae-Hwan Cho, 2015. “An impact analysis of FMD news on pork demand in korea.” The Korean Journal of Community Living Science 26(1): 75-85.
- 김양석, 이충권. 2018. “뉴스기사 분석을 통한 사회이슈와 가격에 관한 연구-조류인플루엔자와 달걀 가격 중심으로.” 『스마트미디어저널』 7(1): 45-51.
- 신성호, 이미경, 송사광. 2018. “LSTM 네트워크를 활용한 농산물 가격 예측 모델.” 『한국콘텐츠학회 논문지』 18(11): 416-429.
- 문권순. 1997. “벡터자기회귀(VAR) 모형의 이해.” 『통계분석연구』 2(1): 23-56.
- 이형용, 여민수, 홍승지. 2017. “마늘 도매가격 시계열 예측 모형 비교. FTA 이행에 따른 낙농·유가공 산업의 경제적 영향 분석.” 『농업경제연구』 40(2): 59-63.
- 홍승지, 김성훈, 이을경. 2015. “오차수정모형을 이용한 고추의 재배면적반응함수 추정.” 『농업경영·정책연구』 42(3): 507-522.
- Alibabaei K., Gaspar P. D., Lima T. M. 2021. “Crop Yield Estimation Using Deep Learning Based on Climate Big Data and Irrigation Scheduling.” Energies 14(11): 3004.
- Tran, Q., and Song, S. (2017). “Water level forecasting based on deep learning: A use case of trinity river-Texas-The United States.” The Journal of Korean Institute of Information Scientists and Engineers 44(6): 607-612.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. 2014. “Dropout: a simple way to prevent neural networks from overfitting.” The Journal of Machine Learning Research 15: 1929-1958.
- Hochreiter, S., Schmidhuber, J. 1996. “Lstm can solve hard long time lag problems.” In Proceedings of the 9th International Conference on Neural Information Processing Systems(NIPS'96), MIT Press. Cambridge, MA, USA, 473-479.