The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Livestock Price Forecasting using Long Short-Term Memory Units: the Case of African Swine Fever and the COVID-19 Pandemic

Do-il Yoo,
Associate Professor,
Department of Agricultural Economics and Rural Development,
Seoul National University,
scydl8@snu.ac.kr

Selected Paper prepared for presentation at the 2022 Agricultural & Applied Economics Association Annual Meeting, Anaheim, CA; July 31-August 2

Copyright 2022 by Do-il Yoo. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Livestock Price Forecasting using Long Short-Term Memory Units: the Case of African Swine Fever and the COVID-19 Pandemic[1]

Do-il Yoo[2]

1. Introduction

African Swine Fever (ASF) is a contagious virus disease observed at for any pig farm. It brings about lots of socioeconomic problems threatening the food security of all countries and also affects the imports of pigs from neighboring countries. For example, due to ASF outbreaks in China, pork imports in Korea decreased by 14.3-21.4%. As a result, pork price in Korea has changed rapidly; it rose by 40% in September 2019 and dropped in half just after one month. Such a rapid price volatility involves market disturbance with the demand of excessive social cost. Thus, it would be beneficial to provide more accurate price prediction as both farmers and consumers can effectively react to market changes induced by ASF outbreak.

For this study, we pay attention to recent methodologies based on deep learning (DL) algorithm, which is a specific concept of machine learning, with the understanding that its prediction power is known to be more robust than that of conventional time series models (Adhikari and Agrawal, 2013; Gamboa, 2017). Especially, we employ Long Short-Term Memory (LSTM) units as a key model. LSTM is an advanced model of Recurrent Neural Networks (RNNs), which is one of methods relying on DL (Che et al., 2017). While the

[1] This article is still under revision and may have lots of changes up to the AAEA presentation. Please do not cite this article.

[2] Associate Professor, Department of Agricultural Economics and Rural Development, Seoul National University, scydl8@snul.ac.kr

application of RNNs suffers from a vanishing gradient problem and an exploding gradient problem, LSTM provides alternative ways to mitigate those problems by RNNs. Also, LSTM is known to be efficient in capturing long-term dependencies and inner relations among multiple time series (Hochreiter and Schmiduber, 1997).

Due to its advanced efficiency, LSTM has been applied to various fields such as image recognition, language learning, time series prediction, and so on (Graves et al., 2007; Perez-Oritz et al. 2003; Wierstra et al., 2005). However, LSTM has been rarely used in the field of agriculture, such as agricultural water, farm land, remote sensing data, and so on. (Ndikumana et al., 2018; You et al., 2017; Zhang et al., 2018). Most recently, Jiang et al. (2018) suggested prediction of corn yields over county levels at the Corn Belt in the United States using LSTM, which is closely connected to our research object.

Based on LSTM, we devise a prediction model for pork price with multivariate time series data considering livestock disease outbreaks, such as ASF, Foot Mouth Disease (FMD), and Highly Pathogenic Avian Influenza (HPAI), and substitution relation such as beef and broiler. We gather weekly-based time series data from January 2010 through April 2022 covering all outbreaks of ASF, FMD, and HPAI. In particular, we consider the COVID-19 pandemic broke out in December, 2019. As an atypical data, we consider text queries concerning those outbreaks, which are gathered from on-line websites such as news, blogs, or social media services.

Prediction algorithm is as follows: first, we normalize all time-series data using min-max scaling. Second, we generate both training and testing datasets. Third, we generate LSTM networks using AdamOptimizer provided by TensorFlow with Python as an optimization function. Finally, we predict pork price and compare results with the mean absolute percentage error (MAPE). In addition, conventional time series models like Vector Error Correction Model (VECM) can be considered as a benchmark case for comparison. We expect our study can provide some introductory implication in using LSTM for the field of agricultural economics and for the livestock policy makers in case livestock diseases break out.

2. Model

As a benchmark, we use Threshold Vector Autoregressive (TVAR) model and Threshold Vector Error Correction Model (TVECM) to identify the nonlinear structure of the livestock product prices caused by the infectious diseases in livestock. Both models rely on the threshold model, an analytical method dividing a sample into two or more regions according to the threshold values when the subject has a nonlinear relationship (Zapata and Gauthier, 2003).

TVAR model is based on the Vector Autoregressive (VAR) model with threshold effects. A single TVAR model with one threshold value is shown in the following equation (1) (Hansen, 2000).

$$Y_t = \begin{cases} c_1 + \beta_{1j} X_t + \epsilon_{1t} & if \;\; q_t < \gamma_i \\ c_2 + \beta_{2j} X_t + \epsilon_{2t} & if \;\; \gamma_i \le q_t \end{cases} \tag{1}$$

where, $X_t = \left[Y_{t-1}\, Y_{t-2}\, Y_{t-3} \cdots \, Y_{t-p}\right], \quad Y_t = [P_1\, P_2\, P_3]$

Y_t is a dependent variable vector, and q_tT is a threshold time series variable. The VAR model can be divided into two or more regimes based on γ_i which is the threshold value. This is summarized in the following equation (2) to (3).

$$Y_t = (c_1 + \beta_1 X_t + \epsilon_{1t}) I(q_t < \gamma_i) + (c_2 + \beta_2 X_t + \epsilon_{2t}) I(q_t \ge \gamma_i) \tag{2}$$

$$Y_t = C + \delta X_t(\gamma) + e_t \tag{3}$$

where. $X_t(\gamma) = [X_t I(q_t < \gamma_i)\ \ X_t I(q_t \geq \gamma_i)]', \quad \delta = [\beta_1\ \beta_2]$

I is an indicator function that has a value of 1 if the condition is satisfied and a value of 0 otherwise. The TVAR model has a nonlinear / discontinuous form because it has different parameter according to the indicator function I, such as β_1 and β_2. Therefore, the appropriate parameter estimation could be derived from the sequential conditional least squares as equation (4) and the threshold value can be estimated the least squares as shown in equation (5) (Hansen, 1997).

$$\hat{\delta}(\gamma) = (\textstyle\sum_{t=1}^{n} X_t(\gamma)X_t(\gamma)')^{-1}(\textstyle\sum_{t=1}^{n} X_t(\gamma)Y_t) \tag{4}$$

$$\hat{e_t}(\gamma) = Y_t - \hat{\delta}(\gamma)X_t(\gamma)$$

$$\widehat{\sigma_n}^2(\gamma) = \frac{1}{n}\sum_{1}^{n} \hat{e_t}(\gamma)^2$$

$$\hat{\gamma} = \text{argmin}_{\gamma \in [\overline{\gamma}\, \underline{\gamma}]}\ \widehat{\sigma_n}^2(\gamma) \tag{5}$$

In order for the estimation results to be meaningful, the threshold effects must exist. For this reason, it is necessary to test the threshold effect. However, under the null hypothesis, there is a problem that the threshold value is not identified in testing the threshold effect. As shown in the equation (6), this can be solved through F- test for residuals both under the null hypothesis and under the alternative hypothesis by assuming that the error term e_t follows iid(independent identically distributed). This is because if e_t follows iid, it is possible to approximate for asymptotic distribution through bootstrapping (Hansen, 1997).

$$F_n = n\left(\frac{\widetilde{\delta_n^2}-\widehat{\delta_n^2}}{\widehat{\delta_n^2}}\right),\ \widetilde{\delta_n^2}\ \text{is sum of squar residuals under alternative hypothesis} \tag{6}$$

$$F_n(\gamma) = n\left(\frac{\widetilde{\delta_n^2}-\widehat{\delta_n^2}(\gamma)}{\widehat{\delta_n^2}(\gamma)}\right)$$

TVECM also has a form in which the threshold effect is added to the Vector Error Correction Model (VECM). When there exists a cointegrating relationship, VECM assumes that long-run equilibrium relationship is linear, while TVECM assumes that long-run equilibrium is nonlinear. This is because the presence of transection costs and/or fixed adjustment costs may prevent economic agents from correcting the error continuously (Balke and Formby, 1997). The basic VECM is shown the following equation (7).

$$\begin{bmatrix}\Delta P_t^1\\ \Delta P_t^2\\ \Delta P_t^3\end{bmatrix} = \begin{bmatrix}C_1\\ C_2\\ C_3\end{bmatrix} + \sum_{i=1}^{n}\begin{bmatrix}\beta_i^{11} & \beta_i^{12} & \beta_i^{13}\\ \beta_i^{21} & \beta_i^{22} & \beta_i^{23}\\ \beta_i^{31} & \beta_i^{32} & \beta_i^{33}\end{bmatrix}\begin{bmatrix}\Delta P_{t-i}^1\\ \Delta P_{t-i}^2\\ \Delta P_{t-i}^3\end{bmatrix} + \begin{bmatrix}\lambda^1\\ \lambda^2\\ \lambda^3\end{bmatrix}[ECT_{t-1}] + \begin{bmatrix}\epsilon_t^1\\ \epsilon_t^2\\ \epsilon_t^3\end{bmatrix} \tag{7}$$

In equation (7), β_i is the short-term price change, λ is the adjustment speed to the long-term average, and ECT_{t-1} is the error correction term. Balke and Formby (1997) extended the VECM by applying the concept of threshold cointegration, which means TVECM. It is assumed that there exists the threshold in ECT, so that the model is a suitable for explaining short- and long-term changes in variables with nonlinear long-run equilibrium. TVECM implies that ECT can be divided into several regions according to the threshold values, so that each region has different adjustment speeds as shown in equation (8) to (9).

$$\begin{bmatrix} \Delta P_t^1 \\ \Delta P_t^2 \\ \Delta P_t^3 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} \beta_i^{11} & \beta_i^{12} & \beta_i^{13} \\ \beta_i^{21} & \beta_i^{22} & \beta_i^{23} \\ \beta_i^{31} & \beta_i^{32} & \beta_i^{33} \end{bmatrix} \begin{bmatrix} \Delta P_{t-i}^1 \\ \Delta P_{t-i}^2 \\ \Delta P_{t-i}^3 \end{bmatrix} + \begin{bmatrix} \lambda_1^1 \\ \lambda_1^2 \\ \lambda_1^3 \end{bmatrix} [ECT_{t-1}] + \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \\ \epsilon_t^3 \end{bmatrix} \tag{8}$$

$$if \;\; ECT_{t-1} < \gamma$$

$$\begin{bmatrix} \Delta P_t^1 \\ \Delta P_t^2 \\ \Delta P_t^3 \end{bmatrix} = \begin{bmatrix} C_4 \\ C_5 \\ C_6 \end{bmatrix} + \sum_{i=1}^{n} \begin{bmatrix} b_i^{11} & b_i^{12} & b_i^{13} \\ b_i^{21} & b_i^{22} & b_i^{23} \\ b_i^{31} & b_i^{32} & b_i^{33} \end{bmatrix} \begin{bmatrix} \Delta P_{t-i}^1 \\ \Delta P_{t-i}^2 \\ \Delta P_{t-i}^3 \end{bmatrix} + \begin{bmatrix} \lambda_2^1 \\ \lambda_2^2 \\ \lambda_2^3 \end{bmatrix} [ECT_{t-1}] + \begin{bmatrix} e_t^1 \\ e_t^2 \\ e_t^3 \end{bmatrix} \tag{9}$$

$$if \;\; ECT_{t-1} \geq \gamma$$

It shows that there exists only one threshold in the ECT, so that the model is divided into two regions. In the model, the short-term effects are classified into β_i and b_i, and the adjustment coefficient to long-run equilibrium is also divided into λ_1 and λ_2.

Also, we devise a Long Short-Term Memory (LSTM) model. In an RNN, tanh is used as an activation function to train the model in a non-linear way. However, there is a long-term dependency problem caused by a "vanishing gradient" problem in the RNN's BPTT, in which the gradient (weights update rate) disappears as the value (derivative value of the tanh function with respect to h_t) less than 1 continues to multiply. Thus, the state of a relatively distant past time point has almost no effect on an output of the present time point. As a result, the model relies only on short-term data and has a limit in achieving the best performance. To solve this problem, Hochreiter et al. (1997) suggested the LSTM model.

** Figure 1 **

Figure 1 shows the internal structure of LSTM and its process. LSTM is the model in which forgetting and memory (f_t), the input (i_t), the inner cell state candidate ($\widetilde{C_t}$), the conveying and inner cell state at time point t (C_t), and the output (o_t) are added to the RNN model. Especially, C_t, which penetrates all time points, greatly contributes to solving the long-term dependency problem. The order of each part and the internal algorithm can be explained by the following process:

$$f_t = \sigma(W_{xh(f)}x_t + W_{hh(f)}h_{t-1} + b_{h(f)}) \tag{10}$$

$$i_t = \sigma(W_{xh(i)}x_t + W_{hh(i)}h_{t-1} + b_{h(i)}) \tag{11}$$

$$\widetilde{C_t} = tanh(W_{xh(\widetilde{C_t})}x_t + W_{hh(\widetilde{C_t})}h_{t-1} + b_{h(\widetilde{C_t})}) \tag{12}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{13}$$

$$o_t = \sigma(W_{xh(o)}x_t + W_{hh(o)}h_{t-1} + b_{h(o)}) \tag{14}$$

$$h_t = o_t \odot tanh(C_t) \tag{15}$$

$$*\odot = Hadamard\ product, \sigma = sigmoid\ function = \frac{1}{1 + e^{-x}}$$

Equation (24), output of the forget gate, determines whether the historical state is forgotten by the combination of x_t and h_{t-1}. The output value of this step is converted to a number between 0 and 1 by the sigmoid function and multiplied by C_{t-1} (memory of past data, i.e., historical state) to determine how much past data to preserve or forget. A value of 0 indicates forgetfulness, and 1 indicates memorization of past data. Equations (25) and (26) are involved in the storage of the inner cell state of time point t. Equation (25), output of the input gate, determines how much data of time point t are memorized. In other words, it has a value between 0 and 1, indicating the degree of memorizing for the new information. At the same time, Equation (26) generates the inner cell state candidate of time point t. Equation (27) generates the new cell state at time point t and passes it on to the LSTM cell at the next time point (t + 1). In other words, LSTM solves the RNN's long-term dependency problem by adjusting the memorization and forgetfulness of the past and presents the state through Equations (24)–(27). In the end, the output is decided by Equations (28) and (29). Equation (28), output of the output gate, decides which part of the new cell state will become output. A value of the new cell status is converted through the tangent function and calculated with the result value of Equation (28) to produce the final output of time point t, as shown in Equation (29).

3. Data

This study applies TVAR and TVECM to the Korean market in detail using the price of pigs, broilers, and eggs from January 2011 to May 2017 in Korea. Furthermore, we set the infectious diseases data as threshold variables. The price data and the infectious diseases data are obtained from Korea Institute for Animal Products Quality Evaluation (KAPE) and Korea Animal Health Integrated System (KAHIS) in Korea, respectively. The infectious diseases data[3] is used by multiplying the number of infected livestock by the average carcass weight per species. The summary statistics on price data and on diseases are shown in form Table 1 to Table 2.

** Table 1 **

[3] Based on the data in KHAIS, pig infectious diseases include FMD (Foot Mouth Disease), PRRS (Porcine Reproductive and Respiratory Syndrome), CSF (Classical Swine Fever), Aujeszky's disease, and Brucellosis. Broiler and laying hens diseases include fowl typhoid, pullorum disease, Newcastle disease, HPAI (Highly Pathogenic Avian Influenza). In the cases of broiler, Tuberculosis is also added.

** Table 2 **

4. Results

Since the models used in this study rely on time series data, the stationarity of the data is verified through the unit root test. As a result of the ADF test, there exist unit root in all prices. Also, it shows that the first difference data of all prices are stationary. Then, the optimal lag is determined based on AIC, SIC, and HQIC statistics. The optimal lag of all livestock products model is 1 as shown in Table 4.

** Table 3 **

** Table 4 **

Then, we conduct Hansen and Seo (2002) test to test whether there are thresholds in cointegrating relationship considering the threshold cointegration effect established in Balke and Formby (1997). If there are more than one threshold in the error correction term between the distribution channels, which means that long-run equilibrium relationship between the distribution channels has nonlinear structure due to the infectious diseases. In addition, it is necessary to test the threshold cointegration to determine which model is more suitable for analysis among TVAR model and TVECM.

** Table 5 **

The results of threshold cointegraion show that there are no threshold cointegration between the distribution stages' prices of pig. In the case of broiler prices, there are two threshold cointegration both farm to wholesale and wholesale to retail. The result of egg shows that both farm to retail and wholesale to retail are significant. The purpose of this study is to analyze to whether the livestock procducts' prices from farm to retail have a nonlinear structure according to the incidence of infectious diseases, so that TVECM seems to be not appropriate model for analysis because it is not able to grasp the relation of prices of the entire distribution stages. On the other hand, TVAR model can consider the interaction of prices of the entire distribution stages, so we conduct analysis based on TVAR model.

In order to validate threshold model, Hansen (1999) showed that the threshold effects should exist in the VAR model composed of prices for distribution stages as shown in Table 6.

** Table 6 **

The results of the threshold effects test represent that VAR model for prices of distribution stages has two thresholds for infectious diseases, which means that VAR model would be divided into three regimes due to infectious diseases. Considering the above, the results of TVAR with two thresholds are represented in Table 7 to Table 9.

** Table 7 **

The threshold values for pig TVAR model in Table 7 are estimated to 812,1278 kg and 1,278.733 kg. In regime 1 where the pig diseases are less than 812.1278 kg, constant and retail price are reject the null hypothesis. Specifically, the retail lag price has a negative impact on the farm price and retail price, also it has negative impact on the wholesale price in regime 1. All of the lag price is significant in regime 2, where the pig diseases are between 812.1278 kg and 1,278.733 kg. The farm and retail lag prices have negative impacts on all prices, and the wholesale lag price has a positive impact on all prices. In regime 3 where the pig diseases are over 812.1278 kg, the retail lag price has a negative impact on farm and wholesale price, and has a positive impact on retail price.

The threshold values for broiler TVAR model in Table 8 are estimated to 3,600 kg and 8016 kg. The farm lag price has a positive impact on all prices in regime 1 where the broiler diseases are less than 3,600 kg. In the regime 2, the wholesale lag price affects positively on all prices, and besides the farm lag price has a negative effect on retail price. In the regime of highest incidence of diseases for broiler, only the wholesale lag price is statistically significant.

Table 9 shows the results of egg TVAR model with two thresholds. The threshold values are estimated to 1,477.827 kg and 166,029 kg, respectively. The farm and wholesale lag prices affect positively on farm price in regime 1. In regime 2 and regime 3, only the retail lag price is statistically significant for all prices. Especially, the impacts of retail lag price for farm and wholesale price are positive in regime 2, but the impact are negative in regime 3. This is interpreted that if diseases are less than a certain value, the retail lag price has a positive impact on the farm and wholesale price, but if diseases exceeds a certain value, then the retail lag price affects negatively on those prices.

5. Conclusion

This study used the TVAR model and TVECM to determine whether the prices of pig, broiler, and egg have a nonlinear structure due to the incidence of livestock infectious diseases. Prior to the analysis, we conducted unit root test to ensure the stability of the data. The results showed that all prices of pig, broiler, and egg are nonstationary. Thereafter, threshold cointegration test is performed, which showed that there is no threshold cointegration in the prices of pig. Moreover, the prices of all livestock products have the threshold effect. The threshold values of the pig were estimated to 812.1278kg and 1,278.733kg. The threshold values of broiler were estimated to 3,600kg and 8,016kg, and threshold values of egg were estimated to 1,477.827kg and 166,029kg. Then, Granger causality was conducted to test the causality for the livestock products' price of different distribution stages, after then we analyzed the generalized impulse response function through causality path. As a result of analyzed the impulse response function, it was confirmed that the shock in each distribution stages is different according to the incidence of infectious diseases. The shock of the prices has different duration and speed according to the separated regimes by diseases.

This suggests that if infectious diseases occur on a large-scale, government policies should be implemented to suit with each livestock products. For instance, in the case of broiler, the shock of prices tends to maintain about 3 to 6 weeks when diseases occur on a large-scale, so that it is necessary to implement short-term stabilization policy, such as the release of government' and private' stockpiles. In the case of pig, the shock of price lasts more than 12 weeks when a large-scale diseases occur. Therefore, it is necessary not only short-term policies such as securing supply through the import, but also long-term policies such as the supply and demand forecasting system to maintain the appropriate number of pig or the promotion for consumption of substitution goods through discount. Finally, in the case of egg, the shock of the price lasts more than about 12 weeks as with pig. Consequently, it is necessary to implement long-term policies such as restoring infrastructure for laying hens, along with short-term policies which increase supplies such as stockpile and imported egg.

In conclusion, the significant results of this study are expected to be useful resources for the price stabilization policy and distribution policy for livestock products. However, this study has the limitation in that it does not take into consideration the various economic factors related to the price of livestock products. It is also expected that more detailed interpretation will be made if economic analysis using specific numerical values such as variance decomposition analysis considering economic causality is added.

References

Adhikari, R. and R. Agrawal, 2013, *An Introductory Study on Time Series Modeling and Forecasting*. Germany: Lap Lambert Academic Publishing.

Che, Z., P. Sanjay, K. Cho, D. Sontag and Y. Liu, 2017, “Recurrent Neural Networks for Multivariate Time Series with Missing Values,” ICLR.

Gamboa, J., 2017, “Deep Learning for Time-Series Analysis,” *arXiv preprint arXiv:* 1701.01887.

Graves, A., S. Fernández, M. Liwicki, H. Bunke, and J. Schmidhuber, 2007, “Unconstrained Online Handwriting Recognition with Recurrent Neural Networks,” Proceedings of the 20th International Conference on Neural Information Processing Systems. NIPS 2007. USA: Curran Associates Inc., 577–584.

Hochreiter, S. and J. Schmidhuber, 1997, “Long Short-Term Memory,” *Neural Computation*, 9(8): 1735-1780.

Jiang, Z., C. Liu, N. P. Hendricks, B. Ganapathysubramanian, D. J. Hayes, and S. Sarkar, 2018, “Predicting County Level Corn Yields using Deep Long Short Term Memory Models,” *arXiv preprint arXiv*:1805.12044.

Ndikumana, E., D. H. T. Minh, N. Baghdadi, D. Courault, and L. Hossard, 2018, “Deep Recurrent Neural Network for Agricultural Classification using Multitemporal SAR Sentinel-1 for Camargue, France,” *Remote Sensing*.

Perez-Ortiz, J. A., F. A. Gers, D. Eck, and J. Schmidhuber, 2003, “Kalman Filters Improve LSTM Network Performance in Problems Unsolvable by Traditional Recurrent Nets,” *Neural Networks*, 16(2): 241–250.

Wierstra, D., J. Schmidhuber, and F. J. Gomez, 2005, "Evolino: Hybrid Neuroevolution/Optimal Linear Search for Sequence Learning," Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh, 853–858.

You., J., X. Li, M. Low, D. Lobell, and S. Ermon, 2017, "Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data," *In AAAI*: 4559-4566.

Zhang, J., Y. Zhu, X. Zhang, M. Ye, and J. Yang, 2018, "Developing a Long Short-Term Memory (LSTM) based Model for Predicting Water Table Depth in Agricultural Areas," *Journal of Hydrology*, 561: 918-929.

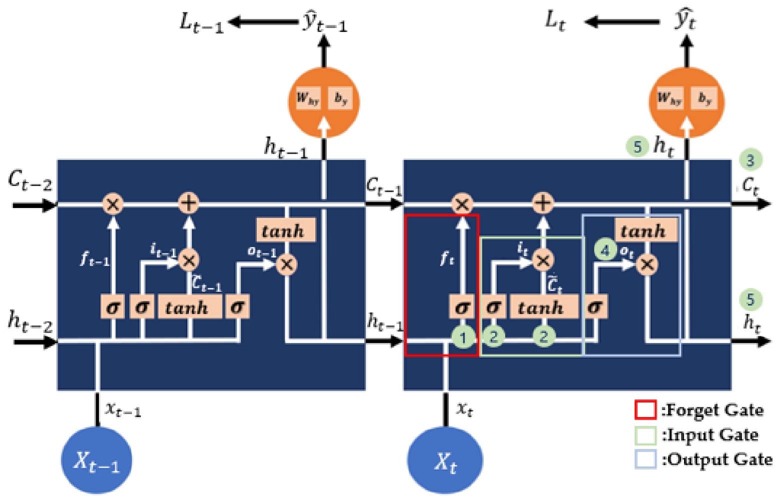

Figure 1. Internal structure of long short−term memory (LSTM).

Table 1. Price summary statistics from 2011 to 2017 (KRW/kg)

	Price	Mean	Std.	Min	Max
Pig	Farm	3,182.50	614.46	1,839	5,170
	wholesale	4,625.77	893.11	2,673	7,515
	Retail	19,023.88	2,521.98	12,214	24,950
Broiler	Farm	1,660.64	316.59	1,048	2,546
	wholesale	3,155.51	431.70	2,196	4,231
	Retail	5,626.39	488.79	4,644	7,123
Egg	Farm	2,024.44	392.27	1,344	3,409
	wholesale	2,228.62	385.64	1,490	3,628
	Retail	2,996.21	406.13	2,285	4,808

Table 2. Disease summary statistics from 2011 to 2017

	Obs.	Occurrences	Mean(kg)	Std. (kg)	Min(kg)	Max(kg)
Pig disease	215	157,814	48,730.97	284,193.8	112.49	3,836,720
Broiler disease	184	2,736,037	11,379.47	46,490.41	3.11	455,003.3
Laying hens disease	103	21,495,268	91,692.49	656,212.5	1.56	7,944,253

Table 3. Results of Augmented Dickey-Fuller test for unit root

Type	Price		ADF test	
			t-stat.	p-value
Pig	Level	Farm	-0.8709	0.3382
		wholesale	-0.8625	0.3418
		Retail	-0.3023	0.5764
	D(-1)	Farm	-16.211	0.0000***
		wholesale	-16.406	0.0000***
		Retail	-15.681	0.0000***
Broiler	Level	Farm	-1.6742	0.0890*
		wholesale	-0.9904	0.2884
		Retail	-0.6584	0.4315
	D(-1)	Farm	-24.6403	0.0000***
		wholesale	-11.3081	0.0000***
		Retail	-12.5793	0.0000***
Egg	Level	Farm	-1.0386	0.2694
		wholesale	-0.9654	0.2986
		Retail	-0.6391	0.4400
	D(-1)	Farm	-14.5967	0.0000***
		wholesale	-13.0746	0.0000***
		Retail	-18.8839	0.0000***

***: $p < 0.01$, **: $p < 0.05$, *: $p < 0.1$

Table 4. Optimal lag selection

	AIC	SIC	HQIC	Optimal lag
Pig	price (-2) (41.4909)	price (-1) (41.6428)	price (-1) (41.5641)	1
Broiler	price (-7) (43.0791)	Price (-1) (43.3421)	Price (-1) (43.2635)	1
Egg	price (-3) (38.3212)	price (-1) (38.5158)	price (-2) (38.4138)	1

a) The blanket means statistics

Table 5. Results for Hansen and Seo(2002) cointeration test

	Null hypothesis	Farm to wholesale	wholesale to Retail	Farm to Retail
Pig	t-statistics	4.9407	14.6855	14.4052
	P-value	0.18	0.10	0.10
Broiler	t-statistics	38.7673	32.1415	13.8768
	P-value	0.00***	0.00***	0.24
Egg	t-statistics	15.1331	24.1565	26.7018
	P-value	0.32	0.00***	0.00***

a) ***: $p < 0.01$

Table 6. Results for threshold effects test

		LR-statistics	P-value
Pig	Linear VAR vs. 1 threshold VAR	28.98	0.12
	Linear VAR vs. 2 threshold VAR	100.92	0.02**
Broiler	Linear VAR vs. 1 threshold VAR	44.17	0.01**
	Linear VAR vs. 2 threshold VAR	101.92	0.03**
Egg	Linear VAR vs. 1 threshold VAR	70.01	0.00***
	Linear VAR vs. 2 threshold VAR	114.57	0.00***

a) ***: $p < 0.01$, **: $p < 0.05$

b) Bootstrapping repeats 100 times

Table 7. Results of TVAR model for pig

	Farm price	wholesale price	Retail price
regime 1	Pig diseases ≦ 812.1278kg		
Farm price(-1)	0.5382 (0.3895)	0.7895 (0.5643)	1.7205 (2.8837)
wholesale price(-1)	0.2956 (0.2695)	0.4243 (0.3905)	-0.0684 (1.9955)
Retail price(-1)	-0.0291*** (0.0076)	-0.0413*** (0.0110)	0.3864*** (0.0560)
c	655.7839*** (101.2629)	932.3193*** (146.7106)	6,451.8378*** (749.7696)
regime 2	812.1278kg < Pig diseases ≦ 1,278.733kg		
Farm price(-1)	-406.8953*** (65.7822)	-591.6832*** (95.3058)	-3,103.2220*** (487.0636)
wholesale price(-1)	280.9183*** (45.2660)	408.4941** (65.5817)	2138.0734*** (335.1575)
Retail price(-1)	-0.1207*** (0.0201)	-0.1755*** (0.0292)	-0.5419*** (0.1490)
c	974.9594** (454.9076)	1,418.2307** (659.0737)	15,938.3320*** (3368.2200)
regime 3	Pig diseases > 1,278.733kg		
Farm price(-1)	-10.0156 (23.7271)	-14.4539 (34.3760)	112.2430 (175.6799)
wholesale price(-1)	7.6053 (16.3222)	10.9828 (23.6477)	-76.7311 (120.8523)
Retail price(-1)	-0.0338** (0.0166)	-0.0490** (0.0240)	0.8278*** (0.1229)
c	496.2407** (211.3994)	720.1629** (306.2771)	873.2963 (1565.2401)

a) ***: $p < 0.01$, **: $p < 0.05$

b) The blanket means standard error