

Predicting Korea' business-cycle regimes using OnBid auction data

Jin Gi Kim

POSTECH, Pohang, Republic of Korea

Hyun-Tak Lee

Korea Asset Management Corporation, Busan, Republic of Korea, and

Bong-Gyu Jang

POSTECH, Pohang, Republic of Korea

116

Received 1 November 2020

Revised 28 January 2021

20 February 2021

Accepted 21 February 2021

Abstract

Purpose – This paper examines whether the successful bid rate of the OnBid public auction, published by Korea Asset Management Corporation, can identify and forecast the Korea business-cycle expansion and contraction regimes characterized by the OECD reference turning points. We use logistic regression and support vector machine in performing the OECD regime classification and predicting three-month-ahead regime. We find that the OnBid auction rate conveys important information for detecting the coincident and future regimes because this information might be closely related to deleveraging regarding default on debt obligations. This finding suggests that corporate managers and investors could use the auction information to gauge the regime position in their decision-making. This research has an academic significance that reveals the relationship between the auction market and the business-cycle regimes.

Keywords Machine learning, Business cycles, Deleveraging, OnBid, Public auction

Paper type Research paper

1. Introduction

A public auction refers to an auction conducted by the public sector such as the government, local governments or public agencies. It is usually carried out when the public sector requests disposal of assets to Korea Asset Management Corporation (KAMCO) to forcibly collect unpaid taxes. In contrast, an auction takes place in general when financial firms or creditors request forcible sale to the court to liquidate insolvent debts. KAMCO is contributing to the national finances by carrying out public auctions through OnBid, which



is an online platform that publicly sells or leases assets such as those owned by the public sector and seized for the purpose of forcible sale [1].

This paper explores whether the average of the successful bid rate of the OnBid public auction (hereafter, the OnBid auction rate in short), collected by foreclosures or public-owned assets, is capable of identifying and predicting the Korea business-cycle expansion and contraction regimes now or three months later. Nowadays, household, corporate and government debts are on the rise, possibly because low interest rates and large-scale quantitative easing have been implemented worldwide because of COVID-19. When interest rates rise someday in the future to prevent the economy from overheating, the excessively accumulated debt for a while will likely worsen the debt repayment ability of households and companies. Doing so can cause events of default on debts such as non-payments of principals, interests and taxes in arrears to increase, which triggers deleveraging (debt reduction) in a way of the sale of real estate collateralized assets in the public auction market. To make matters worse, the asset value may plunge further in a self-reinforcement pattern, signaling economic contractions. Therefore, we hypothesize that the OnBid auction rate, which is the representative of the public auction market, can be indicative of the severity of contraction just before deleveraging begins or at the time of recession.

To test the hypothesis mentioned earlier, this paper conducts logistic regression and support vector machine for the identification and prediction of the coincident and three-month-ahead regimes by using the OECD based Recession Indicators for Korea. We find that the estimated probability of contraction now and after three months increases as the OnBid auction rate decreases in a way that the bid price decreases compared to the appraised value. This finding arises because if deleveraging begins at the brink of economic contractions, the supply of the auction assets on fire sale increases, whereas the demand of buying these assets decreases, inducing the bid price to fall further relative to the appraised value.

One implication is that business managers or investors can use the OnBid auction rate as an important decision-making indicator when diagnosing the business-cycle regimes. For example, the fact that the OnBid auction rate starts to decline below the long-term average as deleveraging starts gives a sign of economic contraction, giving a hint of managing risk. In contrast, when the OnBid auction rate starts to rise from the bottom, this sign of economic recovery can encourage them to increase investment.

Market participants usually have a great interest in asset markets such as stocks, bonds, real estate and raw materials in terms of profit seeking and risk hedging. However, their interest in the public auction market is relatively low. Hence, this paper deserves to elucidate the importance of the public auction market, which plays a role of recovering economic health through deleveraging.

The structure of this paper is as follows: Section 2 examines the main differences between conventional studies and our research, Section 3 explains the research methods, Section 4 describes the data, and Section 5 shows the results of empirical analysis. Section 6 discusses the implications, and Section 7 makes concluding remarks.

2. Literature review

One branch of conventional studies regarding the auction and public auction markets assumes that these markets lag behind the business-cycle regimes, and thus attempts to predict the trend of the successful bid rates by using macroeconomic variables. [Kim and Park \(2013\)](#) use a vector error correction model (VECM) to estimate the explanatory power for the apartment auction rate of macroeconomic variables. [Baek and Jeong \(2015\)](#) analyze the effect of macroeconomic variables by using the VECM. Conversely, our focal point is on

whether the public auction rate can be used as coincident and leading indicators to predict the regimes.

Another branch of the studies related to economic prediction mainly focuses on forecasting stock and option market movements (Yoon and Kim, 2014; Kim, 2016; Sim, 2016) or finding leading-indicator variables that can predict the regimes. Ivanova *et al.* (2000) show that the German government bond spread (9–10 year bond yields minus 1–2 year bond yields) can predict the German regimes. Ahrens (2002) argues that the long- and short-term spreads of major countries in the world (the USA, Japan, Germany, The Netherlands, Canada, France and the UK) can be used as a leading-indicator variable. Furthermore, Moneta (2005) demonstrates that interest rate spreads (10-year bond yields minus three-month LIBOR rates) predict the European Union's regimes. Chauvet (1999) claims that the S&P 500 index helps to forecast the regimes. Gilchrist and Zakrajšek (2012) show that the credit spread is a good leading-indicator variable. We attempt to perform empirical analysis by adding the OnBid auction rate to the major variables discussed above.

Besides, we seek to carry out a multi-period binary classification in a cross-sectional manner by making a number of building blocks, each of which has one binary dependent variable (i.e. the OECD recession indicator) and a set of independent variables as one unit of analysis. As an example, Birchenhall *et al.* (1999) and Birchenhall *et al.* (2001) evaluate the predictive power of the USA and UK regimes by the using multi-period logistic approach. Shumway (2001) also uses this method to predict the probability of default estimated based on market and financial data.

The aforementioned studies on economic prediction generally deal with continuous variables such as GDP growth rate based on an ARIMA series. Doing so might capture the trend of GDP growth, but the predicted growth level itself is unclear for judging the exact regime position. For this reason, we want to use the OECD recession indicator as a binary dependent variable.

Moreover, the conventional studies mentioned earlier typically assume a linear relationship between dependent and independent variables. Recently, numerous studies challenge to use machine learning that can consider nonlinear relationships. Lin and Pai (2010) and Hung and Lin (2013) attempt to predict the business cycles by using support vector regression. Gogas *et al.* (2015) use support vector machine to determine the presence of economic recession and show that that support vector machine provides better results than do logistic and probit regression approaches. Lee (2017) combines a deep learning algorithm with a technical indicator to predict the Korean KOSPI stock index. Bae and Yoo (2018) predict the real transaction price index of apartment sales by using machine learning (support vector machine, random forest, gradient boosting regression tree, deep neural network and long short-term memory [LSTM]); and time-series approaches (autoregressive moving average model, vector auto-regression model and Bayesian vector auto-regression model) and then conclude that machine learning outperforms and time series approaches. Tang *et al.* (2020) contend that the stock price index prediction using the LSTM artificial neural network algorithm helps to forecast the business cycles.

As such, using machine learning has the advantage of superior predictive performance than the traditional linear analysis, but it has difficulty in interpreting the results that are also vulnerable to overfitting. We try to overcome the shortcomings of machine learning by using nested cross-validation and bootstrap methods introduced in Section 4, and then compare these results of support vector machine with those of linear logistic regression[2].

3. Research method

3.1 Logistic regression analysis

Logistic regression is a statistical model used to predict the probability of an event using a linear combination of independent variables. Currently, the logistic regression analysis is

widely used for classification and prediction in various fields such as medicine and communication.

The logistic regression model ensures that the value of a dependent variable is always within the range of 0 to 1 because of the transformation of odds and logit, regardless of the numerical size of the independent variables. The odds transformation is expressed as the ratio for the probability of success and failure, and the calculation formula is as follows:

$$\text{odds} = \frac{p(y = 1|x)}{1 - p(y = 1|x)}.$$

The logit transformation is a function that takes the logarithm of odds, and when the domain is $[0, 1]$, co-domain is $(-\infty, \infty)$. The formula is as follow:

$$\text{logit} = \log \frac{p(y = 1|x)}{1 - p(y = 1|x)}.$$

The logistic regression model is one special case of the ordinary linear regression model and has a form:

$$\text{logit} = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n = B \cdot X,$$

where B and X on the right side are expressed in a vector form. The dependent variable having a value between 0 and 1 is converted into a value between $(-\infty, \infty)$ through odds and logit transformation. In addition, the transformed value is predicted by using a linear regression model. That is, the odds and logit transformation of a dependent variable can be expressed as a linear function of independent variables:

$$\text{logit}(p) = \log \frac{p(y = 1|x)}{1 - p(y = 1|x)} = B \cdot X.$$

The goal of logistic regression is to calculate the probability that a dependent variable is equal to 1 or belongs to a specific class, given independent variable:

$$P(y = 1|X) = \frac{e^{B \cdot X}}{1 + e^{B \cdot X}}.$$

3.2 Support vector machine

Support vector machine, one of the popular machine-learning methods, aims to scrutinize classification rules for given data. Currently, support vector machine is also being used to solve various pattern recognition problems such as medicine, text and image classification.

Support vector machine intends to find a hyperplane that classifies two categories, for example. When a set of independent variables can linearly separate a dependent variable, the plane separating this set is called a hyperplane. It can be expressed as follows:

$$w^T X - b = 0.$$

Here, the vector w is a normal vector, which is perpendicular to the hyperplane. In fact, a number of hyperplanes exist, but support vector machine is designed to select a hyperplane

that maximizes the distance between the hyperplane and the nearest independent variables. The support vector of a given hyperplane is defined as follows:

X^+ : The closest independent variable to hyperplane among the independent variables corresponding to $y = 1$.

X^- : The closest independent variable to hyperplane among the independent variables corresponding to $y = -1$.

The hyperplane passing X^+ while having the same normal vector refers to a plus hyperplane and the hyperplane passing through X^- refers to a minus hyperplane. The plus hyperplane and the minus hyperplane can be represented as follows.

$$w^T X^+ - b = 1, \quad w^T X^- - b = -1. \quad (1)$$

Notably, the distance between the plus and minus hyperplanes is defined as a margin, and the relationship between the margin λ and the hyperplane can be expressed as follows:

$$X^+ = X^- + \lambda w. \quad (2)$$

Equation (2) means that the plus hyperplane is λ away from independent variables toward the normal vector. Equations (1) and (2) are used to compute the margin λ :

$$w^T X - b = 1,$$

$$w^T (X^- + \lambda w) - b = 1,$$

$$w^T X^- - b + \lambda w^T w = 1,$$

$$-1 + \lambda w^T w = 1,$$

$$\lambda = \frac{2}{w^T w}.$$

Note that support vector machine aims to maximize the margin:

$$\max \frac{2}{w^T w} \iff \min \frac{w^T w}{2} \quad \text{such that } y_i(x^T X_i - b) \geq 1.$$

Here, the constraint above means that no independent variable exists between the plus hyperplane and the minus hyperplane, which reflects the assumption that the support vector (X^+ , X^-) is the closest independent variable to the hyperplane in each domain.

Sometimes, no hyperplane can classify both categories. The first solution is to modify the optimization formula such that there are different categories of independent variables between the margins, called C-support vector machine. The second solution is to select a nonlinear classification boundary, called Kernel-support vector machine.

Specifically, C-support vector machine allows different classification of independent variables within the margin but imposes a penalty. Independent variables belonging to either the plus (minus) hyperplane impose a penalty for distance ζ beyond the plus (minus) hyperplane. The optimization rule of C-support vector machine has:

$$\min \frac{w^T w}{2} + C \sum_{i=1}^n \zeta_i \quad \text{such. that} \quad y_i(w^T X_i - b) \geq 1 - \zeta_i.$$

Note that as the penalty C increases, the width of the margin decreases.

Figure 1 illustrates the idea of Kernel-support vector machine. As seen earlier, linear classification is impossible to separate two classes, and thus each independent variable is mapped as a high-dimensional space through function ϕ to find a hyperplane. We will use both C-support vector machine and Kernel-support vector machine by selecting optimal penalty size and kernel in which the accuracy and AUC [3] are the highest for each data.

4. Data

4.1 Variable selection

The dependent variable is the OECD based Recession Indicators for Korea from the Period following the Peak through the Trough between January 2004 and September 2019 (189 months), provided by the US Federal Reserve Board (FRB). Note that economic contraction has one, whereas economic expansion has zero on a monthly basis (Figure 2).

The independent variables are the OnBid auction rate (whole country, Seoul, residential and non-residential), CD interest rate, bond spread (10-year government bond yield minus one-year government bond yield), credit spread (Corporate bond BBB – Corporate bond AA) and annual KOSPI return (Table 1).

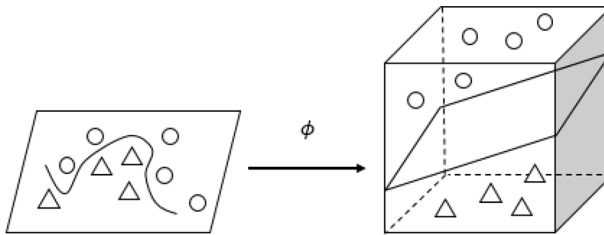


Figure 1. Kernel-support vector machine concept

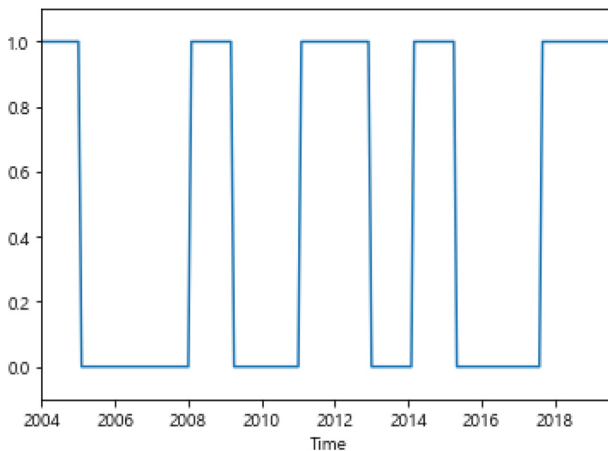


Figure 2. OECD based recession indicator

	Whole country	Seoul	Residential	Non-residential
<i>Panel A: OnBid auction rate (Unit: %)</i>				
Mean	75.24	73.11	78.55	57.65
Variance	33.33	81.86	37.16	53.07
Maximum	92.21	104.65	93.52	81.83
Median	75.03	73.12	79.76	57.12
Minimum	57.21	36.84	61.56	34.43
<i>Panel B: Other leading-indicator variables (Unit: %)</i>				
	CD rate	Bond spread	Credit spread	KOSPI return
Mean	3.02	0.75	5.13	9.37
Variance	1.57	0.35	1.61	409.70
Maximum	6.03	2.88	6.27	64.36
Median	2.79	0.64	5.77	6.71
Minimum	1.34	0.03	2.14	-46.09

Notes: This table reports descriptive statistics of independent variables used in this analysis. The OnBid auction rate (the successful bid rate of the OnBid public auction) is available from the Onbid webpage (www.onbid.co.kr/op/cma/gnrdatamn/generalDataList.do). The other four variables known as leading-indicator variables mentioned in Section 2 can be downloaded from the Economic Statistics System of the Bank of Korea (<https://ecos.bok.or.kr/>)

Table 1.
Descriptive statistics

We then perform multi-period binary classification by using these dependent and independent variables. For instance, coincident analysis carries 189 building blocks, each of which has a pair of independent and dependent variables this month. In a similar manner, three-month-ahead leading analysis creates 186 building blocks, each of which has three-month-lagged independent variables and the dependent variable this month.

4.2 Data pre-processing

We take three main pre-processing steps before implementing empirical analysis.

First, winsorization is performed to mitigate the effect of outliers located in the lower 5% and upper 95% percentile of each independent variable. Second, standardization after the winsorization is performed to make the independent variables to mean 0 and standard deviation 1. The main reason for standardization is that estimated parameters can be significantly affected by scaling when performing regularization imposing penalty terms. For example, the variance of the OnBid auction rate (whole country) before standardization is about 33%, while the variance of the CD rate is only about 1.6%. Therefore, penalizing these variables can generate a scale effect, but this effect can be offset in advance as the variance of the two variables is forced to be one as a result of the standardization. Third, nested cross-validation is performed to estimate hyperparameters [4]. Cross-validation can be available because the building blocks composed of dependent and independent variables are arranged in a cross-section manner. In addition, we perform *stratified* cross-validation that keeps the frequency of contractions (89 out of 189 months; about 47%) consistent.

Figure 3 describes the concept of the nested cross-validation. First, we divide the 189 building blocks of the coincident analysis, for example, into an outer loop and an inner loop, while each loop keeps a frequency of about 47% economic contractions. The purpose of the inner loop is to estimate the optimal hyperparameter for each outer

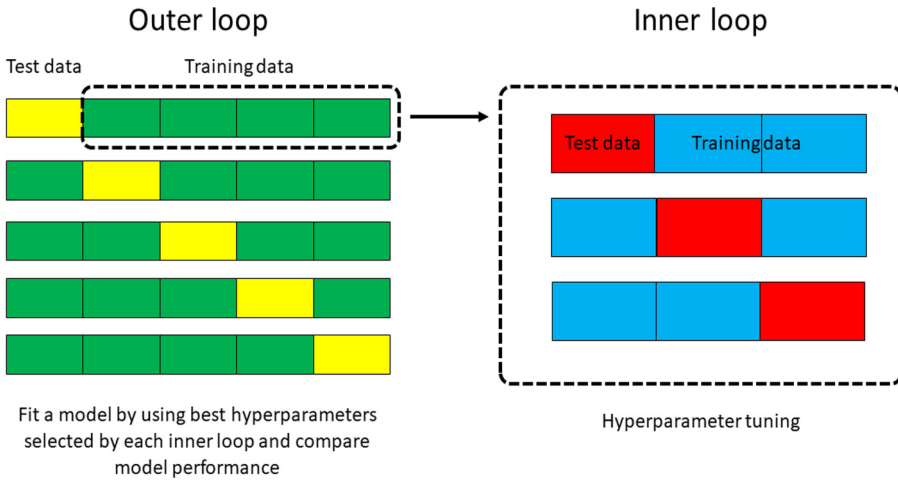


Figure 3.
Concept of nested cross-validation

loop. The purpose of the outer loop is to evaluate the model’s performance through the remaining test data for validation. One merit of performing the nested cross-validation is that we can reduce the problem of sample selection bias. Specifically, we divide the outer loop into five parts, and the inner loop into three parts to perform the nested cross-validation. By repeating this for five outer loops, we can finally choose the optimal hyperparameters that yield the best performance.

4.3 SHapley Additive exPlanation

SHapley Additive exPlanation (SHAP) refers to a marginal effect (marginal contribution) to calculate the importance for one characteristic based on game theory (Lundberg and Lee, 2017). Recently, the application of machine learning and deep learning tends to become complex, making it difficult to interpret the results. With the advancement of computer technology, it has become possible to numerically analyze the marginal effect by evaluating the presence or absence of the independent variable of interest on the model results (e.g. probability of contraction). If SHAP is 0, it means that the independent variable has no impact on economic contractions. If SHAP is greater than 0, the variable increases the probability of contraction [5].

One great merit of using SHAP is that it allows for estimating an interaction effect between independent variables, whereas the traditional approaches do not typically. For example, the traditional regression analysis assumes that the other independent variables remain unchanged when interpreting the coefficient of a particular variable of interest. Therefore, one often depends on pre-processing to alleviate multicollinearity in advance before going into empirical analysis. For this reason, the traditional regression analysis has difficulty in considering the interaction effect in principle. Suppose that two building blocks have the same OnBid auction rate. The fact that logistic regression has the same coefficient indicates that the effect on the probability of contraction should be the same as each other although the two blocks have a distinct combination of the other variables. In the meantime, numerical analysis using SHAP can overcome this drawback discussed earlier. For instance, if the credit spread of the first building block is larger than that of the second building block, the effect of the first building block’s OnBid auction rate is likely to be greater

than that of the second building block, which can be amplified by the interaction effect.

5. Empirical analysis

We reproduce the regime classification and predict the three-month-ahead regimes by using a method of bootstrap. First, 80% of the total building blocks (hereinafter, training data) is randomly selected by allowing for repetition after applying the optimal hyperparameters derived from the nested cross-validation and then trained for each model. Similarly, we analyze the remaining 20% (hereinafter, test data) for comparison later. Overall, we perform 1,000 times of this training and test procedure. This bootstrap method of using random sampling spends a lot of time, but it also has a strength of increasing the result reliability while preventing the problem of sample selection bias. In the rest of the analysis, we focus on the residential bid rate to conserve time and space [6].

5.1 Coincident analysis using logistic regression

We begin by the bootstrap results (excluding intercept) of logistic regression estimated from training data (Table 2). Note that the scale of the independent variables is already adjusted from data pre-processing, and it is thus easy to compare the magnitude and sign of the regression coefficients.

The mean regression coefficient of the OnBid auction rate (-2.19) points out that the lower auction rate increases log odds ratio (log probability of contraction relative to expansion). Further, a distribution of 1,000 coefficients shows that the 25th and 975th percentile values within the 95% confidence interval are estimated to be -3.8 and -0.9, respectively. The fact that this interval [-3.8, -0.9] is away from zero ensures that the result is statistically significant, implying that the OnBid auction rate has useful information for detecting economic contraction today.

One thing to emphasize here is that the OnBid auction rate seems to have the weakest effect, when it comes to the absolute size of the coefficients. As emphasized in subsection 4.3, however, the interpretation of regression coefficients requires special attention. For instance, the coefficient of OnBid ratio (-2.19) is only sensible when the other independent variables remain unchanged. During economic contractions, a lot of variables tend to move in tandem; for example, the government bond spread decreases, while the credit spread increases at the same time. Clearly, such an interaction effect cannot be captured by the traditional regression interpretation.

To overcome this, we attempt to use SHAP along with bootstrap to identify variable importance under interaction. Note that SHAP analysis will interpret the probability of

	OnBid rate	CD rate	Bond spread	Credit spread	KOSPI return	R ²
mean	-2.19	7.49	-3.14	7.11	-2.37	0.14
s.d.	0.772	1.659	0.931	1.660	0.897	0.149
95% CI	[-3.84, -0.86]	[4.47, 11.00]	[-5.27, -1.57]	[4.13, 10.74]	[-4.16, -0.61]	[-0.17, 0.42]

Table 2. Results of logistic coefficients using bootstrap (training data)

Notes: This table reports bootstrap training results of estimated coefficients in logistic regression and R². We run 1,000 times to compute the mean, standard deviation (s.d.), and 95% confidence interval (CI). For example, the lower 2.5% percentile is ranked in 25th among the 1,000 results and the upper 97.5% percentile is ranked in 975th

contraction p in an “absolute” sense, whereas the regression coefficients (Table 2) interpret log odds ratio $\log(p/(1-p))$ in a “relative” sense.

Specifically, SHAP bootstrap is designed in two ways. First, we examine only variable importance in terms of the absolute value of SHAP, as we already explored whether the independent variables have a positive (or negative) effect on economic contractions in Table 2. Similarly, we examine the mean, standard deviation, and 95% confidence level of the absolute SHAP values by performing 1,000 bootstraps. Second, we compare the results of the training data with those of the test data for robustness (Table 3). As a result, we find that the difference between the estimated probability, which arises from the presence and absence of the OnBid auction rate, is estimated to be about 8% (Panel A; Panel B). At the 95% confidence level, it is worth noting that the OnBid auction rate increases the probability by up to about 13%.

One may wonder if the bootstrap results of the training and test data are statistically different. For this purpose, we perform a t -test to analyze whether the means of the estimated results from the two data groups are the same (Panel C). The t -value and p -value are estimated to be 0.327 and 0.744, respectively. These statistics cannot reject the null hypothesis that the means of the two data groups do not differ from each other at the 5% level, demonstrating the effect of the OnBid auction rate on the regime identification.

Now, we turn to estimation performance determined by logistic regression using the entire sample data (Table 4). Excluding the OnBid auction rate (Panel A) allows the model to predict 60 months as in contraction out of 89 months that are actually in contraction between January 2004 and September 2019, indicating a 67% recall rate. Excluding the OnBid auction rate also produces about a 71% precision rate as 60 months actually belong to be in contraction among the 84 predicted months. On the other hand, including the OnBid auction price rate (Panel B) shows that the recall rate is about 76% (=68/89), while the precision is about 78% (=68/87). All predictive metrics of including the OnBid auction rate outperform those of excluding the OnBid auction rate, which are consistent with the bootstrap results.

	OnBid rate	CD rate	Bond spread	Credit spread	KOSPI return
<i>Panel A: 1,000 bootstraps with training data</i>					
mean	0.078	0.199	0.096	0.225	0.078
s.d.	0.025	0.024	0.025	0.028	0.029
95% CI	[0.032,0.127]	[0.147,0.243]	[0.052,0.146]	[0.168,0.279]	[0.017,0.136]
<i>Panel B: 1,000 bootstraps with test data</i>					
mean	0.077	0.198	0.096	0.226	0.078
s.d.	0.024	0.022	0.025	0.027	0.030
95% CI	[0.035,0.126]	[0.154,0.241]	[0.053,0.150]	[0.175,0.281]	[0.018,0.138]
<i>Panel C: T-test for comparison of the means between two groups (t-test for equal mean)</i>					
t -value	0.327	0.716	-0.002	-0.873	-0.120
p -value	0.744	0.474	0.999	0.383	0.904

Notes: This table reports bootstrap results of estimated SHAP in logistic regression. We run 1,000 times to compute the mean, standard deviation (s.d.), and 95% confidence interval (CI) for training (Panel A) and test data (Panel B). Panel C shows t - and p -values of t -test for comparison for the means between training and test data; the null hypothesis is that both means are equal to each other

Table 3. Results of SHAP estimation using bootstrap (logistic regression)

5.2 Coincident analysis using support vector machine

The previous section focuses on the linear relationship between the OECD recession indicators and the independent variables. In contrast, this section aims to make best use of both C-support vector machine imposing a penalty and Kernel-support vector machine using a nonlinear kernel [7]. As a result of the nested cross-validation, we find that Gaussian (radius basis function) kernel yields the best performance in identifying the regime classification. It is informative that Gaussian kernel is expressed by a hyperparameter related to the standard deviation, so it has no coefficients that represent the hyperplane like a linear algorithm. Therefore, we focus on the bootstrap results using SHAP (Table 5).

We find that the OnBid auction rate contributes to about 10% on average for the regime identification and up to about 15% within the 95% confidence interval (Panel A; Panel B). The fact that the *p*-value is about 0.393 demonstrates that the means of

		Prediction		Total
		Expansion	Contraction	
<i>Panel A: Excluding the OnBid auction rate</i>				
Actual	Expansion	76	24	100
	Contraction	29	60	89
	Total	105	84	189
<i>Panel B: Including the OnBid auction rate</i>				
Actual	Expansion	81	19	100
	Contraction	21	68	89
	Total	102	87	189

Table 4.
Confusion matrix for coincident analysis (logistic regression)

Notes: This table reports results of confusion matrix for coincident analysis by using logistic regression. Note that the total number of building blocks used in multiperiod binary classification is 189 months that consist of 89 months in contraction and 100 months in expansion

	OnBid rate	CD rate	Bond spread	Credit spread	KOSPI return
<i>Panel A: 1,000 bootstraps using training data</i>					
mean	0.101	0.150	0.119	0.224	0.063
s.d.	0.026	0.027	0.026	0.029	0.017
95% CI	[0.051,0.149]	[0.097,0.198]	[0.072,0.173]	[0.167,0.279]	[0.033,0.101]
<i>Panel B: 1,000 bootstraps using test data</i>					
mean	0.100	0.149	0.118	0.225	0.060
s.d.	0.026	0.026	0.024	0.026	0.016
95% CI	[0.054,0.152]	[0.096,0.199]	[0.076,0.169]	[0.169,0.271]	[0.034,0.096]
<i>Panel C: t-test for comparison of the means between two groups (t-test for equal mean)</i>					
<i>t</i> -value	0.854	0.763	0.617	-1.062	3.230
<i>p</i> -value	0.393	0.446	0.538	0.289	0.001

Table 5.
Results of SHAP estimation using bootstrap (support vector machine)

Notes: This table reports bootstrap results of estimated SHAP in support vector machine. We run 1,000 times to compute the mean, standard deviation (s.d.), and 95% confidence interval (CI) for training (Panel A) and test data (Panel B). Panel C shows *t*- and *p*-values of t-test for comparison for the means between training and test data; the null hypothesis is that both means are equal to each other

both data groups do not differ from each other (Panel C) against sample selection bias, highlighting the influence of the OnBid auction rate on the regime identification.

In the meantime, the KOSPI return contributes to about 6% on average and up to about 10% within the 95% confidence interval. Importantly, the fact that the p -value turns out to be 0.001 indicates that the means of the two data groups might be different. This result implies that the KOSPI return might suffer from the problem of sample selection bias, and thus does not convey coincident information.

Next, we move on to estimation performance determined by support vector machine using the entire sample data (Table 6). Excluding the OnBid auction rate produces the recall rate to be about 80% (=71/89) and the precision rate to be about 87% (=71/82). On the other hand, including the OnBid auction rate lets the recall rate about 91% (=81/89) and the precision rate 83% (=81/98). In summary, these two metrics draw contrary results because including the OnBid auction rate induces the higher recall rate but the lower precision rate than does excluding the OnBid auction rate.

	Prediction		Total
	Expansion	Contraction	
<i>Panel A: Excluding the OnBid auction rate</i>			
Actual			
Expansion	89	11	100
Contraction	18	71	89
Total	107	82	189
<i>Panel B: Including the OnBid auction rate</i>			
Actual			
Expansion	83	17	100
Contraction	18	81	89
Total	101	98	189

Table 6. Confusion matrix for coincident analysis (support vector machine)

Notes: This table reports results of confusion matrix for coincident analysis by using support vector machine along with the entire data. Note that the total number of building blocks used in multiperiod binary classification is 189 months that consist of 89 months in contraction and 100 months in expansion

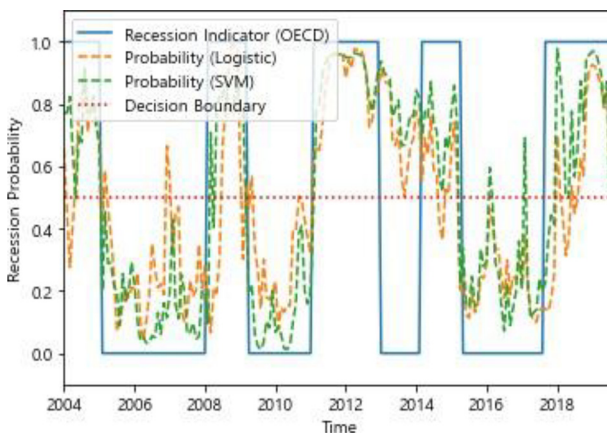


Figure 4. Comparison of the probability of contraction for coincident analysis

To reconcile these contrary results, we introduce f1-score that combines the two metrics together: $f1\text{-score} = 2 \times (\text{recall} \times \text{precision}) / (\text{recall} + \text{precision})$. As a result, we find that the f1-score excluding the OnBid auction rate is about 83%, whereas the f1-score including the OnBid auction rate is about 87%. This result supports the idea that the OnBid auction rate can serve as detecting the current regime.

To help understand the estimated results, Figure 4 illustrates the predicted probability of contraction estimated by both approaches along with the OECD-based recession indicators. Note that when the predicted probability is over 0.5, this signal means that the current regime is in economic contraction. Otherwise, it can be interpreted as economic expansion.

5.3 Three-month-ahead leading analysis

So far, we have explored whether logistic regression and support vector machine can identify the current regime. One may wonder whether the OnBid auction rate provides leading information in predicting the future regime, and if it can, how many months it can lead. To solve this conundrum, this section examines whether the OnBid can predict three-month-ahead regime [8].

First, we investigate whether each independent variable increases or decreases the three-month-ahead probability of contraction by using logistic regression (Table 7). Commonly, how the independent variables work looks almost similar to the results of the coincident

Table 7.
Results of logistic coefficient estimation using bootstrap (training data)

	OnBid rate	CD rate	Bond spread	Credit spread	KOSPI return	R^2
mean	-0.70	8.58	-3.53	8.30	-0.52	-0.01
s.d.	0.60	1.74	0.78	1.67	0.59	0.17
95% CI	[-1.97,0.35]	[5.44,12.26]	[-5.19,-2.08]	[5.39,12.03]	[-1.78,0.52]	[-0.36,0.32]

Notes: This table reports bootstrap training results of estimated coefficients in logistic regression and R^2 . We run 1,000 times to compute the mean, standard deviation (s.d.), and 95% confidence interval (CI). For example, the lower 2.5% percentile is ranked in 25th among the 1,000 results and the upper 97.5% percentile is ranked in 975th

Table 8.
Results of SHAP estimation using bootstrap (logistic regression and support vector machine)

	OnBid rate	CD rate	Bond spread	Credit spread	KOSPI return
<i>Panel A: 1,000 bootstraps using logistic regression and test data</i>					
mean	0.027	0.211	0.107	0.246	0.020
s.d.	0.019	0.022	0.017	0.024	0.017
95% CI	[0.000,0.067]	[0.169,0.254]	[0.074,0.140]	[0.200,0.294]	[0.000,0.060]
<i>Panel B: 1,000 bootstraps using support vector machine and test data</i>					
mean	0.071	0.189	0.135	0.253	0.056
s.d.	0.026	0.023	0.026	0.025	0.015
95% CI	[0.032,0.130]	[0.141,0.230]	[0.085,0.185]	[0.205,0.301]	[0.032,0.091]

Notes: This table reports SHAP bootstrap results of logistic regression and support vector machine by using the 20% test data of the 186 building blocks. We run 1,000 times to compute the mean, standard deviation (s.d.) and 95% confidence interval (CI) for logistic regression (Panel A) and support vector machine (Panel B)

analysis (Table 2): a lower OnBid auction rate is associated with a higher probability of economic contraction in three months.

One different thing is that the variable importance regarding interest rates such as the CD rate, bond spread, and credit spread increases, whereas the variable importance of the OnBid auction rate and KOSPI return decreases. For example, the mean coefficient (-0.70) of the OnBid auction rate is lower than that of the coincident analysis (-2.19). Even, the 95% confidence interval includes zero and the mean of R^2 becomes negative, implying that the leading effect of the OnBid auction rate might diminish [9].

Next, we compare the variable-importance results of logistic regression and support vector machine by using SHAP bootstrap along with the test data (Table 8). Logistic regression (Panel A) shows that the average contribution of the OnBid auction rate in three months is estimated to be about 3%, although the 95% confidence interval still involves a statistical significance baseline of zero. However, support vector machine (Panel B) shows that the average contribution rate is about 7% and up to 13% within the 95% confidence interval.

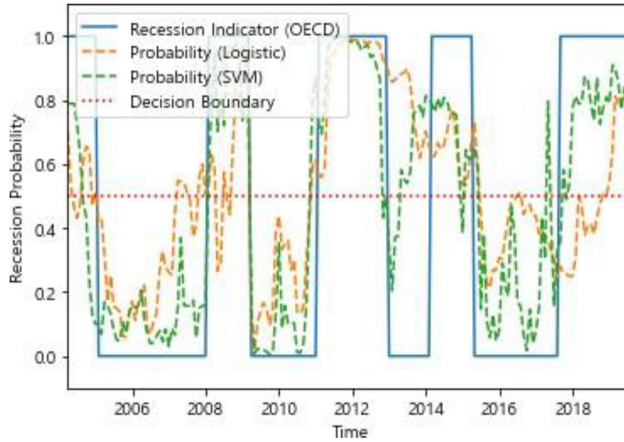
Now, we turn to prediction performance (Table 9). Excluding the OnBid auction rate using logistic regression (Panel A) produces a recall rate of 74% (=64/86) and a precision rate of 70% (=64/92). In contrast, including the OnBid auction rate using logistic regression (Panel B) shows that the recall rate is about 73% (=63/86) and the

	Prediction		
	Expansion	Contraction	Total
<i>Panel A: Excluding the OnBid auction rate (Logistic regression)</i>			
Actual			
Expansion	72	28	100
Contraction	22	64	86
Total	94	92	186
<i>Panel B: Including the OnBid auction rate (Logistic regression)</i>			
Actual			
Expansion	73	27	100
Contraction	23	63	86
Total	96	90	186
<i>Panel C: Excluding the OnBid auction rate (Support vector machine)</i>			
Actual			
Expansion	82	18	100
Contraction	11	75	86
Total	93	93	186
<i>Panel D: Including the OnBid auction rate (Support vector machine)</i>			
Actual			
Expansion	83	17	100
Contraction	7	79	86
Total	90	96	186

Table 9. Confusion matrix for leading analysis (logistic regression and support vector machine)

Notes: This table reports three-month-ahead prediction results by using logistic regression (Panels A and B) and support vector machine (Panels C and D). Note that the total number of three-month-ahead building blocks used in multiperiod binary classification is 186 months that consist of 86 months in contraction and 100 months in expansion

Figure 5.
Comparison of the
probability of
contraction for
leading analysis



precision rate is about 70% (=63/90). These two results are almost the same as each other, so it seems that the OnBid auction rate has no leading information.

However, support vector machine says the other way around. Excluding the OnBid auction rate (Panel C) yields a recall rate of 87% (=75/86) and a precision rate of 81% (=75/93), whereas including the OnBid auction rate (Panel D) has a recall rate of about 92% (=79/86) and a precision rate of about 83% (=79/96). The first evidence is that these metrics are greater than those of using logistic regression (Panels A and B), and the second evidence is that those of excluding the OnBid auction rate falls behind those of including the OnBid auction rate. Figure 5 plots the predicted probability produced by both approaches.

6. Implications

We present three main findings as below.

First, the credit spread is the most important coincident and leading indicator, which is followed by the CD rate and the bond spread. The first reason why the credit spread increases during economic contraction is clear because companies with low credit ratings must pay higher interest rates as the probability of default increases. Second, the CD rate usually serves as a discount rate that lowers the face value of negotiable deposit certificates, especially in the event of recession. Third, economic contraction lets the bond spread decrease, so even an inverse yield curve rarely happens during economic downturn. Our results are consistent with the intuition mentioned above: the higher credit spread and CD rate are associated with the higher probability and the lower bond spread is associated with the higher probability (Table 2).

Second, the lower OnBid auction rate, defined as the bid price divided by the appraised value, is related to a higher probability of contraction. This finding can be justifiable because supply increases during economic contraction, as collateral assets tend to be on fire sale in the auction market as a consequence of deleveraging, while demand decreases conversely. Therefore, business managers who pay special attention to the business cycles or investors who allocate their assets over time need to monitor closely not only interest rate information such as the credit and bond spreads, but also public auction information as a deleveraging signal. For example, when the bid rate starts to decrease from the long-term average, one can prepare for

an adverse situation where economic contraction might begin. Conversely, a gradual rise in the rates can give the sign of economic recovery.

Third, it turns out that the OnBid auction rate as a representative of the auction market conveys greater regime information than does the KOSPI return as a representative of the stock market. In fact, COVID-19 causes a gap between the real economy and asset markets such as real estate and stock markets to diverge, which is boosted by low interest rates and quantitative easing. We contend that this divergence is not irrelevant to the low predictive power of the KOSPI return, while the public auction market can play an important role in reducing the gap through deleveraging.

7. Concluding remarks

This study explores whether the OnBid auction rate can identify and forecast the business-cycle regimes by using logistic regression and support vector machine along with SHAP bootstrap. We show that the lower OnBid auction rate is closely related to a higher probability of contraction now and three months later. The simple reason comes from the principle of supply and demand because economic contractions usually let the demand decrease but the supply increase. Notably, the OnBid auction rate can also be used as one of the representative indicators of deleveraging; this is why the auction rate is capable of predicting the current and future regimes.

This paper sheds new light on the link between the public auction market and the business-cycle regimes. Indeed, the public auction market can play an important role in reducing debts. Diagnosing the regimes during the COVID-19 era using public auction data might deserve future work.

Notes

1. OnBid, an abbreviation of “online bidding,” is the first public online auction system in Korea (www.onbid.co.kr).
2. We implemented a number of machine-learning algorithms such as Ensemble Tree (e.g. Random Forest, Gradient Boosting Tree) and Multi-Layer Perception (unreported). We find out that as complexity increases along with a number of hyperparameters, model performance in training data set tends to increase, but the performance in test data set tends to decrease, which is a signal of overfitting. Among several algorithms, we select support vector machine that has low overfitting results.
3. AUC refers to “Area under the ROC curve.” A receiver operating characteristic (ROC) curve is a figure that exhibits the diagnostic ability of a binary classification problem. For example, as the AUC score is close to one, it means that the model performance becomes better. For details, please refer to Google Developers: <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc?hl=en>.
4. For reference, the main hyperparameters of logistic regression include penalty series (Lasso and Ridge) and penalty strength, while those of support vector machine are kernel type (Linear, RBF and Polynomial) and penalty strength.
5. Specific details on how SHAP works are beyond the scope of this paper. We want to refer interested readers to one great article to help better understand SHAP: <https://towardsdatascience.com/shap-explained-the-way-i-wish-someone-explained-it-to-me-ab81cc69ef30>.
6. The results using the other auction rates such as Seoul and non-residential are almost similar to those reported in this paper (unreported).

7. The Python package “scikit-learn” provides four kernels in implementing SVM: Linear; Radius Basis Function (RBF); Polynomial; and Sigmoid. We use Grid Search to estimate optimal penalty parameter “C” and kernel function. For details on the kernel functions, interested readers are referred to <https://scikit-learn.org/stable/modules/svm.html#kernel-functions>.
8. Using Grid Search, we find that the RBF kernel is also optimal in the leading analysis.
9. Note that R^2 can be negative when the selected model fits worse than a horizontal line.

References

- Ahrens, R. (2002), “Predicting regressions with interest rate spreads: a multicountry regime-switching analysis”, *Journal of International Money and Finance*, Vol. 21 No. 4, pp. 519-537.
- Bae, S.W. and Yoo, J.S. (2018), “Real estate price index prediction using machine learning method and time series analysis model”, *Korean Association for Housing Policy Studies*, Vol. 26 No. 1, pp. 107-133.
- Baek, S.G. and Jeong, J.H. (2015), “A study on the effects of macroeconomic variables on the public auction market”, *Korea Real Estate Academy Review*, Vol. 62 No. 2, pp. 19-32.
- Birchenhall, C., H., Jessen, H., Osborn, D. and Simpson, P. (1999), “Predicting US business-cycle regimes”, *Journal of Business and Economic Statistics*, Vol. 17 No. 3, pp. 313-323.
- Birchenhall, C., Osborn, D. and Sensier, M. (2001), “Predicting UK business cycle regimes”, *Scottish Journal of Political Economy*, Vol. 48 No. 2, pp. 179-195.
- Chauvet, M. (1999), “Stock market fluctuations and the business cycle”, *Journal of Economic and Social Measurement*, Vol. 25 Nos 3/4, pp. 235-257.
- Gilchrist, S. and Zakrajšek, E. (2012), “Credit spreads and business cycle fluctuations”, *American Economic Review*, Vol. 102 No. 4, pp. 1692-1720.
- Gogas, P., Papadimitriou, T., Matthaiou, M. and Chrysanthidou, E. (2015), “Yield curve and recession forecasting in a machine learning framework”, *Computational Economics*, Vol. 45 No. 4, pp. 635-645.
- Hung, K.C. and Lin, K.P. (2013), “Long-term business cycle forecasting through a potential intuitionistic fuzzy least-squares support vector regression approach”, *Information Sciences*, Vol. 224, pp. 37-48.
- Ivanova, D., Lahiri, K. and Seitz, F. (2000), “Interest rate spreads as predictors of German inflation and business cycles”, *International Journal of Forecasting*, Vol. 16 No. 1, pp. 39-58, available at: www.sciencedirect.com/science/article/abs/pii/S0169207099000291
- Kim, S. (2016), “On the usefulness of risk-neutral skewness and kurtosis for forecasting the higher moments of stock returns”, *Journal of Derivatives and Quantitative Studies*, Vol. 24 No. 2, pp. 185-220.
- Kim, S.S. and Park, C.S. (2013), “A study on the effects of macroeconomic variables on apartment auction rate”, *Residential Environment*, Vol. 11 No. 3, pp. 237-249.
- Lee, W.S. (2017), “Forecasting the direction of the Korean KOSPI stock index using deep learning analysis and technical analysis indicators”, *Journal of the Korean Data and Information Science Society*, Vol. 28 No. 2, pp. 287-295.
- Lin, K. and Pai, P. (2010), “A fuzzy support vector regression model for business cycle predictions”, *Expert Systems with Applications*, Vol. 37 No. 7, pp. 5430-5435, available at: www.sciencedirect.com/science/article/abs/pii/S0957417410001107
- Lundberg, S.M. and Lee, S.I. (2017), “A unified approach to interpreting model predictions”, *Advances in Neural Information Processing Systems*, pp. 4765-4774.
- Moneta, F. (2005), “Does the yield spread predict recessions in the euro area?”, *International Finance*, Vol. 8 No. 2, pp. 263-301.
- Shumway, T. (2001), “Forecasting bankruptcy more accurately: a simple hazard model”, *The Journal of Business*, Vol. 74 No. 1, pp. 101-124.

-
- Sim, M. (2016), "Realized skewness and the return predictability", *Journal of Derivatives and Quantitative Studies*, Vol. 24 No. 1, pp. 119-152.
- Tang, Y.M., Chau, K.Y., Li, W. and Wan, T.W. (2020), "Forecasting economic recession through share price in the logistics industry with artificial intelligence (AI)", *Computation*, Vol. 8 No. 3, p. 70.
- Yoon, S.J. and Kim, J.S. (2014), "Leading and following variance risk premiums: evidence from S&P500 and KOSPI200 options", *Journal of Derivatives and Quantitative Studies*, Vol. 22 No. 1, pp. 45-70.

Corresponding author

Hyun-Tak Lee can be contacted at: leeht@kamco.or.kr or lht1107@gmail.com