Macroeconomic and Financial Market Analyses and Predictions with Deep Learning*

Soohyon Kim†

May 29, 2020

Preliminary

Abstract

Since Hinton, Osindero, and Teh (2006) developed the fast learning algorithm, deep learning have been powerful tools that have recently achieved impressive performance on a wide spectrum of industries as well as in academia. For the macroeconomic and financial variables, however, more elaborate approaches need to be taken due to the unique latent features of them. In this regards, we propose novel approaches to apply deep learning to the predictions of time series variables in those fields. Specifically, we suggest ensembles of neural networks and Bayesian learning to estimate the posterior distributions of the forecasting outcomes as the out-of-sample forecasts. The examples are provided with the predictions of monthly Korea's nominal exports and daily Korean won-US dollar exchange rates.

JEL classification: C45, F17, F31

Keywords: machine learning; deep learning; Bayesian neural networks; ensemble learning; uncertainty

*I thank the participants at the seminars of the Bank of Korea for helpful comments and suggestions. The views are those of the author and do not necessarily reflect those of the Bank of Korea. The usual disclaimers apply.

†soohyonkim@bok.or.kr, Bank of Korea,55 Namdaemun-ro, Jung-gu, Seoul, 45310, South Korea.

1. Introduction

Machine learning is quickly being familiar to us as the high-performance algorithms showing outstanding outcomes in many fields of industries and academia. Machine learning combines elements from computational statistics, mathematical optimization, pattern recognition, and predictive analytics (Chakraborty and Joseph, 2017) by which the algorithms train themselves through "learning" the latent pattern underlying in data. As deep learning, more advanced algorithms as a branch of artificial neural networks, being introduced, and more-detailed data is being available, the machine learning approach stretch out to more variety of tasks. However, we are hardly acknowledged that machine (deep) learning is being familiar with in the fields of economics [1], when statisticians have accepted this revolutionary approach as a part of their methodologies(Athey and Imbens, 2019). For decades, economics and finance have heavily relied on econometric models for empirical analyses, but we witness nothing but somewhat meaningful progress as being grasped by "inverting a covariance matrix."(Lopez de Prado, 2018a) While empirical analyses in economics and finance are adhere to such conventional approaches, machine learning is proving itself as a most possible alternative to econometrics especially when it is difficult to obtain the answers to the questions through conventional approaches. Machine learning is a set of algorithms that we can set them to train themselves to have desirable predictive powers for the future outcomes. Even with its lack in the structural backgrounds based on economic theories, we can benefit from this predictive power as long as a research object is concerned with it.

In an environment with overflowing granular data or as known as "big data"[2], we can ask why economists are not eager to quench by employing such a revolutionary technique. Among others, there is a reason why. Unlike data used in the other fields, there are specific characteristics in the data in economics. In economics, especially macroeconomics, we have series updated once in a month or even a quarter so that the length of the data hard to be more than several hundreds or so. The length can be even shorter even less than one hundred for many emerging economies where the histories of those series are not long enough. The amount of data with such lengths is not enough for algorithms to find latent patterns to show the performances that they do in other fields like automated driving, and it could be even worse than the conventional econometric models.[3]

The former studies in economics employing machine learning approach used to be rare to

[1] Recently, there are significant number of applications in finance since 2018, but we can only find literature with machine learning approach in economics quite recently around the beginning of 2020

[2] Big data is bigger in volume, contains more information in a wider range of formats (e.g. text), and is updated more frequently(Chakraborty and Joseph, 2017).

[3] Ho (2019) compares the performances of various approaches in econometrics and machine learning with various setups. See Figure 2

find. Recently, however, there are some frontiers emerging as applications of machine learning approach to economic analysis. (Figure 1) As of the end of January 2020, eight out of ten top downloaded (for the last 60 days) concerning econometrics and statistical methods are the adaptation of machine learning to economics, including natural language processing as known as text mining[4]. In contrast, only one out of ten papers on all time top downloaded papers list in the same category is machine learning related, which is not even an article but the first chapter of a machine learning textbook Lopez de Prado (2018b). This contrary well represents the recent trends in econometric approaches. Currently, machine learning is on the spotlight, and it is gradually turning into a new trend in econometrics. Although, papers applying machine learning to macroeconomic analyses and predictions are still quite rare.

Among the initiative studies on machine learning, this paper try to answer to the following questions. Can we adopt machine learning, especially deep learning, to macroeconomic analyses and predictions? If we can, what is the benefit or how would deep learning make distinction of itself among other approaches in econometrics. We may answer to these questions by putting machine learning on macroeconomic data and compare the results with those from some conventional econometric approaches using the same data set.

The rest of the paper is structured as follows. Section 2 gives overviews of machine learning approach including deep learning. Section 3 characterizes specific features of macroeconomic and financial market variables and finds deep learning approaches applicable to analyses and predictions of macroeconomic and financial market variables. Section 4 shows the examples of deep learning approaches applying to Korean custom clearance exports and Korean won/US dollar exchange rates. Section 5 summarizes our finding and discusses future research avenues.

2. Methodological Frameworks

In this section, we provide basic ideas of the methodological approaches employed in this study. Athey and Imbens (2019) point out that statistical or econometric models and machine learning approaches are different fundamentally even though some basic machine learning algorithms are originated from statistical models such as the logistic regression. Indeed, many models employed in machine learning have been developed decades ago; lasso, ridge regressions are basic machine learning tools[5] However, the fundamental difference between the conventional approaches in economics and finance and machine learning with the

[4]Natural language processing is a branch of machine learning exploring text data.

[5]Although, those models are semi-parametric approaches in statistics, they are not new to econometricians either.

same tools is that the former focus on explaining independent with dependent variables and the latter is on prediction and decision. In econometrics (as much as in statistics), the classical assumptions are important in applying a model to data sets. The main purpose of using logistic models, for example, is to estimate the probabilities of events with dependent variables (Cox, 1958) in statistics. All the statistical significance of parameters and models are important as well. We use AIC (Akaike Information Criterion), SIC (Schwartz Criterion) or likelihood ratios to test the model significance when we implement analyses with logistic models. In machine learning, however, those aspects are hard to be found. When we use the logistic model in machine learning approach, such model significance tests are not the point of interest. In contrast, how good can the model predict the category of an event according to the explicit and implicit features of it. We check the validation other than significance of the algorithm whether it can correctly predict the category through accuracy, precision, recall rate calculated by confusion matrices.[6] The algorithm will eventually train (similar to "estimation" in econometrics) the models, in a way more close to non-parametric estimation in statistics, iteratively to improve the predictive power of the algorithm.

2.1. Machine Learning

Machine learning is development tools for artificial intelligence, in which various mathematical and statistical methods are employed for specific tasks. Machine learning is closely related to computational statistics, which focuses on finding useful patterns from data to make predictions using computers, which calls analytics. We can find various applications of the mathematical optimization theory in the field of machine learning.

As any empirical model approach, machine learning starts with collecting data sets, x and y. We divide data sets into training, validation, and test sets for unique purpose of each set. Training sets are used for training (fitting in statistics) models and we check the models' predictive power with validation sets. Once it turned out to be that no more improvement is possible for a model than training ends and we can exercise prediction for test sets. We can call x as input and the algorithm will compare outputs and y in supervised learning or x in self-supervised learning such as autoencoder[7]. x can include lagged variables for itself and y as well in time series data sets.

$$
\begin{aligned}
x &= x_1, x_2, ..., x_N \\
y &= y_1, y_2, ..., y_N \\
\mathcal{D} &= (x_1, y_1), (x_2, y_2)...(x_N, y_N)
\end{aligned}
$$

[6]Confusion matrix shows right and wrong answers predicted by machine learning algorithm.

[7]There is no explicit y for unsupervised learning.

We call $\mathcal{D}$ as a database and the goal of the machine learning is to find the best model that will predict the future outcome y_{new}. We can define functions for output f_θ and loss $L(f_\theta(x), y)$ with various metrics. The loss functions measure differences between the model output($f_\theta(x)$) and true values (y). Finding the optimal parameter sets for the function (f_θ), which is called learning or training, is to find θ^* satisfying the equation below

$$\theta^* = \underset{\theta \in \Theta}{argmin}\ L(f_\theta(x), y)$$

Loss functions can be determined in terms of training sets of data but also of validation sets.[8] We can find θ to minimize the distance between output and part of y that we use as validation set.[9]

We employ gradient descent, a first-order iterative optimization algorithm to find the local minimum of loss functions, as we use in many statistical packages such as R to optimize an objective function. With given data set, the loss function is now a function of parameter. If a new candidate of parameters $\theta_1 = \theta_0 + \Delta\theta$ satisfies $L(\theta_1) < L(\theta_0)$ then the algorithm will keep searching for new candidates until it reaches to the point $L(\theta_1) = L(\theta_0)$

$$L(\theta + \Delta\theta) \approx L(\theta) + \bigtriangledown L \cdot \Delta\theta$$
$$L(\theta + \Delta\theta) - L(\theta) = \Delta L \approx \bigtriangledown L \cdot \Delta\theta$$

where $\bigtriangledown L$ is a gradient of L in terms of a parameter set. If we let $\Delta\theta = -\eta \bigtriangledown L$, then we obtain the loss function always decreasing at each iteration as below, where η calls the learning rate

$$\Delta L = -\eta \parallel \bigtriangledown L \parallel^2 < 0,\ \eta > 0$$

After we find the optimal parameter sets, we can predict the future outputs of the model with validation sets and then compare the outcome with the realized values. The algorithm can be improved if we repeat this procedure until we reach a certain point before it starts to over-fit.

$$y_{new} = f_\theta(x_{new})$$

[8]Validation set of data is part of whole data (about 10% or so) ,which is used not to train the algorithm but to evaluate and improve the algorithm.

[9]Using validation set to minimize the loss is often called cross-validation.Training and validation proceed repeatedly sometimes showing trade-off relationship between errors and biases.

2.2. *Artificial Neural Networks and Deep Learning*

Machine learning algorithm is basically a linear combination of features (or variables) that we can obtain hyperplane on which vectorized fratures can be projected upon, or by which freatures are categorized. Some features are, however, complex enough to make it hard for a hyperplane to perform proper tasks. In this case, artificial neural networks are the alternative for the models with linearly combine features. Since the first introduction by McCulloch and Pitts (1943), artificial neural networks are not welcomed for decades due to the lack of algorithm by which neural networks can be trained efficiently. After the fast learning algorithm (Hinton et al., 2006) based on backpropagation (Rumelhart, Hinton, and Williams, 1986) has been introduced as deep learning, artificial neural networks became the most preferable platform in machine learning.

2.2.1. *Deep Learning*

The baseline of deep learning framework is quite in line with machine learning algorithm discussed above. We define a sort of neural networks among various application of them such as convolutional neural networks (CNN), recurrent neural networks (RNN) and so forth with parameter $\theta = \{w, b\}$ where w is a set of weights and b is a set of biases. However, the major difference with other machine learning algorithm comes from the loss function that we can define. The narrow choice in loss function is mainly due to two assumptions in the back propagation algorithm by which we can train the networks. The first one is that total loss of the networks over training samples is the sum of loss for each training sample. The second one is that loss for each training example is a function of final output of the networks. [10] There are several viable loss functions for the deep neural networks that satisfy two assumptions, which are the mean squared error (MSE), the cross entropy [11] , and negative maximum likelihood[12] in case we need variational inferences for the networks.

[10]There can be alternative algorithm for the training networks other than backpropagation, however, when most of the deep learning practitioners use popular functional APIs such as Tensorflow, backpropagation is the sole algorithm for the training.

[11]The MSE is standard loss function as

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - f(\theta_j, \mathcal{D}_i))^2$$

In classification or categorization problems, we can use the cross entropy loss function which is

$$CE = [\ln f(\theta_j, \mathcal{D}_i) + (1-y)\ln(1 - f(\theta_j, \mathcal{D}_i))]$$

[12]The negative maximum likelihood is a usual maximum likelihood multiplied by -1. For the negative maximum likelihood loss function to satisfy two assumptions, we actually need one more assumptions which

Due to the first assumption, the loss function for gradient descent in DNN can be described as below.[13]

$$L(\theta_j, \mathcal{D}) = \sum_i^N L(\theta_j, \mathcal{D})/N$$

$$\bigtriangledown L(\theta_j, \mathcal{D}) = \sum_i^N \bigtriangledown L(\theta_j, \mathcal{D})/N$$

We can update parameters by $\theta_{j+1} = \theta_j - \eta \bigtriangledown L(\theta_j, \mathcal{D})$. Our object is to minimize the loss function to a certain point in terms of weights and biases for all layers in the DNN. Unlike classical machine learning, however, the searching process is more complicated and needs much heavier computation. We need more specific routine added to the gradient descent to find the optimal sets of parameters which is the backpropagation algorithm.

Backpropagation The backpropagation algorithm starts from the "back" of neural networks which is the last layer. We can calculate error signals at the last layer and "propagate" those signals to the front layers updating the parameters in neural networks.

$$\delta^L = \sigma'(z^L) \odot \bigtriangledown L(\theta_j, \mathcal{D})$$

where $\odot$ is element wise product (the Hadamard product) and σ' is the first derivative of activation functions, z^L is the inputs for the activation function, and L represents the last layer such that δ^L is the error signal we can get from the last layer of a neural network. Now the error signal for the adjacent layer can be defined as below

$$\delta^{L-1} = \sigma'(z^{L-1}) \odot (w^L)^T \sigma^L$$

We can update the parameters at each layer l in the same spirit as previously described

is i.i.d. (identical independent distribution) which is also a conventional assumption for econometrics. With Gaussian distribution, the negative maximum likelihood is equivalent to MSE criterion and with Bernoulli distribution, it becomes the cross entropy for classification.

[13]Implementing the algorithm need a lot of computer resources to calculate. In order to make it more efficient we can use stochastic gradient descent (SGD) instead as follows,

$$\bigtriangledown L(\theta_j, \mathcal{D}) \approx \sum_i \bigtriangledown L(\theta_j, \mathcal{D})/M, \ M < N$$

where M is batch sizes of whole databases.

gradient descent as below

$$w_j^{l,new} = w_j^l - \eta \bigtriangledown_{w_j^l} L(\theta_j, \mathcal{D})$$
$$b_j^{l,new} = w_j^l - \eta \bigtriangledown_{b_j^l} L(\theta_j, \mathcal{D})$$

Thanks to the backpropagation algorithm we can get the gradient values in terms of weights and biases more efficiently.[14]

$$\bigtriangledown_{w_j^l} L(\theta_j, \mathcal{D}) = \delta^l \, (a^{l-1})^T$$
$$\bigtriangledown_{b_j^l} L(\theta_j, \mathcal{D}) = \delta^l$$

3. Applying Deep Learning to Analyses and Predictions of Macroeconomic and Financial Variables

In this section, we examine the legitimate applications of deep learning to economics. Compared to econometrics, which is already established through tons of theoretical proofs and exercises, deep learning approaches are not technically familiar to us in this field. As formerly being addressed, deep learning can present high potentiality in economics as well. However, unlike data dealt with in engineering such as images, economic data has specific features; low frequency and noise.

3.1. Uncertainty Characterization for Economics Time Series

Due to those characteristics, conventional macroeconomic data may not be relevant for deep learning approach if it is applied to as it is.

In most macroeconomic data, the frequencies that we can obtain are monthly or quarterly, but usually we do not have access to the micro-level data that is used to construct the macroeconomic series. With this shortage in the lengths of the series, there is not enough information that neural network algorithm can exploit to detect the pattern of those series. Therefore, it is uncertain whether weights of neural networks catch the true pattern of series so bear some predictability of series. Otherwise, we often take advantage of high frequency of financial market data. However,

[14] For more intuitive insight of the backpropagation algorithm, visit http://neuralnetworksanddeeplearning.com/chap2.html

Epistemic Uncertainty In a monthly frequency, we only watch one figure for one series. Although, we get the figure once in a month, there must be all the information we need which is summed and contained in that figure but we cannot watch. In this case, we cannot get the true distribution of the events and this is what called the epistemic uncertainty. The epistemic uncertainty can be lessened if we can obtain more data within the monthly figures but those month or quarterly frequencies are already prevail in macroeconomic data. With this shortage in the lengths of the series, there is not enough information that neural network algorithm can exploit to detect the pattern of those series. Therefore, it is uncertain whether weights of neural networks catch the true pattern of series, so that the whole algorithm bears some predictability of the original series. Uncertainty in predictions that arise from the uncertainty in weights is called epistemic uncertainty. Epistemic uncertainty is higher in regions of no or little training data and lower in regions of more training data, so the uncertainty can be reduced if we can get more data.

Aleatoric Uncertainty Financial variables' frequencies can be as high as it can be up to real time tick data, and so we hardly suffer from epistemic uncertainty in dealing with financial data. However, unlike usual big data we use in analyzing with the machine learning or deep learning, financial time series tend to contain significant noise. Market participants formulate expectations of the prices of financial assets but those expectations are hard to be coincides. Therefore, as the frequencies go higher there remain more errors in the data as results of market participants' behavior. That make it hard for neural networks to detect the pattern in the movement of those variables. This is called aleatoric uncertainty and it cannot be improved simply by obtaining more data. Moreover, when there is such uncertainty, the best prediction results we can get is the ones trained to be over-fitted as in Figure 4

3.2. Ensemble Learning

As in statistical mechanics, ensemble learning uses multiple neural networks to obtain better predictive performance than could be obtained from any of the neural networks alone. However, unlike an ensemble method in statistics, which is usually infinite combination of models, an ensemble of neural networks consists of only a finite set of alternative neural networks, although that is able to allow for more flexible structure. Deep learning often finds very suitable hypothesis by which it shows good performance in predictions with a particular tasks. Even though the hypothesis space includes ones that are very well-suited for a task, it may be difficult to find one good hypothesis, especially when data has the characteristics of epistemic uncertainty. Ensemble learning algorithms combine multiple hypotheses, so multiple neural networks, to form a group of hypothesis, by which they represent the epistemic uncertainty in data that is also transcends to the predictions.

Application of ensemble learning There are many optimizing methods to form ensemble of neural networks; bootstraping (bagging), boosting, Bayesian model averaging etc. Those types of ensembles are for the purpose of obtaining the best neural network among them, however, since we try to deal with epistemic uncertainty, which incurs by lack of information, it may be better to leave the uncertainty as a part of the predictions. In this regard, we leave the trained deep neural networks as they are without any selection procedure. Other than selecting one good algorithm among them, we calculate the mean and standard deviation of predictions at each point of periods. With the sets of two parameters we can draw empirical distribution of predicted values by which we can measure the degree of uncertainty and check whether the realized values lie within pseudo confidence intervals.

3.2.1. Variational Inference and Bayesian Neural Networks

Variational Autoencoder Variational autoencoder (VAE), also known as Auto-encoding Variational Bayes, is one of the generative algorithms, consists of an encoder and a decoder, which aims to reconstruct target data from estimated distributions. Encoders return parameters[15] of distributions for latent variables with training data as an input. Decoders turn the latent variables, which are randomly sampled from distributions, into target data. In this regard, latent variables can be seen as a set of control parameters for target data (generated data), which lie on a manifold[16]

The main goal using VAE is to obtain denoised series ($\bar{x}$) out of original series(x) to train deep neural networks. To get $\bar{x}$ we need to have latent variables z which are points on a manifold. However, when x are series full of noise such as daily financial data, there are possibly many manifolds that will represent the target series. In this regard, it had better estimate distribution of latent variables rather than draw out deterministic point estimates. Let say latent variables z follow a certain joint distribution $p(z)$. If there is a deterministic function parameterized by θ such that $g_\theta(\cdot)$, and denoised $\bar{x}$ close to an element x from a target data set $\mathcal{D}$, is random variables generated by $g_\theta(z)$, then $\bar{x}$ is a function of z which

[15] Those parameters are means (μ) and standard deviations (σ) when latent variables follow a Gaussian distribution.

[16] According to the manifold hypothesis, data sets in a high dimensional space can be projected on a manifold, when data points positioning away from a certain manifold are scarce. When there is a manifold that efficiently represents a distribution of a set of data, then Euclidean distances defined on original spaces between data points do not represent the similarity or proximity of them. (Figure 3)

are randomly sampled from $p(z)$.

$$\begin{aligned} &z \sim p(z) \\ &\bar{x} = g_\theta(z) \text{ or} \\ &\bar{x} \sim p(x|g_\theta(z)) = p_\theta(x|z) \end{aligned}$$

$p(z)$ itself is hard to find, but with a target data set x, we can estimate $p(z|x)$ as a posterior distribution of $p(z)$.

$$p(z|x) = \frac{p_\theta(x|z)p(z)}{p(x)}$$

Our goal then can be maximizing $p_\theta(x|z)$ with maximum likelihood estimation (MLE). Unfortunately, we have target data set with full of noise and such high dimensionality, estimation methods using Euclidean distance on original space would not present or approximate true distribution $p_\theta(x|z)$. Since we have no tractable analytical solution for $p(x) = \int_z p(x|z)p(z)dz$ either, in a complex neural networks, we therefore need to approximate the true posterior with well known distribution $q_\phi(z|x)$, also known as variational distribution[17]. (Doersch, 2016)

$$p(z|x) \approx q_\phi(z|x) \sim z$$

Now we want to estimate the parameters for the well known functional form. This can be done by maximizing ELBO (Evidence Lower Bound) or minimizing the Kullback-Leibler divergence[18] between $q_\phi(z|x)$ and the true posterior $p(z|x)$ with respect to ϕ.

$$log(p(x)) = \underbrace{\int log\left(\frac{p(x,z)}{q_\theta(z|x)}\right) q_\theta(z|x)dz}_{\text{ELBO>0}} + \underbrace{KL\left(q_\phi(z|x)||p(z|x)\right)}_{\text{KL divergence>0}},$$

$$\text{where KL divergence} = \int log\left(\frac{q_\theta(z|x)}{p(x|z)}\right) q_\theta(z|x)dz$$

This is also known as the variational free energy (Friston, Mattout, Trujillo-Barreto, Ashburner, and Penny, 2006). The first term is the expected value of the likelihood with respect to the variational distribution. The second term is the Kullback-Leibler divergence between the variational distribution $q_\phi(z|x)$ and the posterior $p(z|x)$. Figure 5 describes the structure

[17]Approximating the true posterior through variational distribution is called as variational inference which is similar to Laplace approximation.

[18]The Kullback-Leibler divergence is a measure of how one probability distribution is different from a second, reference probability distribution. See https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence for more details.

of VAE intuitively.

3.2.2. Conventional Econometric Models for One Period Ahead Prediction

We use two popular approaches to compare those performances with deep learning: vector autoregression (VAR) and vector error correction (VECM) models. For Korean exports prediction models, ADF test reveals that only Korean customs clearance exports has an unit root. We make the variables stationary by year-on-year differences just as the input data for ensemble learning, and use a VAR model for one-month ahead prediction.

For the daily Korean won/US dollar exchange rate prediction model, all daily exchange rates are non-stationary[19] and Johansen cointegration tells us that there are at least one cointegration vector[20]. In this occasion, VECM could be the choice for one-day ahead prediction of Korean won/US dollar exchange rate controlling long-term relationships among exchange rates.

4. Data and Applications

In this section, we provide preliminary and experimental examples of applying deep learning to existing macroeconomic and financial market data. We start with describing data used to train and validate the neural networks and present the outcomes of prediction combined with uncertainty distributions. For the ensemble of neural networks, all time series are used to train and validate the neural networks, and they also used to obtain the prediction with uncertainty distributions. In contrast, all the series are used only to extract the latent vectors for the Bayesian neural network. With the number of independently denoised Korean won/US dollar exchange rates series, the multi-layer perceptron is trained and validated for the outcome.

4.1. Customs Clearance Exports

Customs clearance exports (nominal) is important for a country, such as Korea, which is heavily dependent on external sectors because exports are the main driver of economic growth for such countries. In econometric approach, monthly custom clearance exports data will be used to predict next month of next quarter exports growth so that we can forecast

[19] Brazilian real and Indian rupee against US dollar exchange rates show weak evidence to reject the null hypotheses as in Table **??**

[20] The hypothesis of no cointegration can be rejected with the trace statistic (271.03) compared to the critical critical value at 95% level (232.83) whereas the hypothesis of one cointegration at most cannot with 178.67 compared to 191.81

growth rate eventually. Customs clearance exports data is infamous for its irregularity due to some complicated components affecting the data; nominal exchange rates, business days etc. In this regard, model based predictions of custom clearance export may not be desirable because the latent features are hard to be detected due to the complexity of data. However, the major issue with such monthly data in applying deep learning is inadequate length of series. It therefore turns out to be innate uncertainty as known as epistemic uncertainty. A data set contains epistemic uncertainty where the amount of information is not sufficient to measure the distribution.[21] If we try to apply neural networks to the predictions of Korean custom clearance exports, we confront epistemic uncertainty which makes it hard for deep neural networks to find exact patterns of the data. As a consequence, we will experience somewhat disappointing performance of the networks in predicting future. While such uncertainty cannot be gone, we still can put forward an elaborated approach; ensemble of neural networks. By applying ensemble approach, we do not need to have one exactly trained neural network but hundreds or thousands neural networks trained with the same data[22].

In order to train and predict Korean custom clearance exports, we use historical data of Korean custom clearance exports, University of Michigan Consumer Sentiment Index, and the US Economic Policy Uncertainty Index from January 1978 to July 2019. We preprocessed the data with min-max scaler as in Figure 6. Among many data to which Korean export possibly related, two US data series are used in order to provide more information for the neural networks training[23] The summary statistics is on Table 1.

4.2. Daily Korean Won/US Dollar

For a small open economy with low level of capital control such as Korea, foreign exchange rates are critical financial variables affecting every aspect in the economy. It is therefore important to predict future exchange rates movement and access risks in the foreign exchange market. Foreign exchange rates are market variables materialized by market participants behavior. However, it is forward looking behavior in which market participants trade currencies and futures according to their expectation of the market and the economy. Such behavior depending upon expectation is naturally unpredictable containing lots of error. Sometimes foreign exchange rates, as much as stock prices, show anomalies whenever there are shocks

[21] If we use daily customs clearance exports data, then it would be sufficient amount of information, but there can be other sort of uncertainty such as aleatoric uncertainty. Aleatoric uncertainty will be on the next paragraph.

[22] Mostly, we sample $M < N$ length of data out of series with length N. In this example, however, due to the short length of about $N = 200$, we do not draw samples buy use whole series as it is.

[23] We probably improve the network performance by adding more data series, however, we stop using more series other than these two for the example.

on the market. It is not coincidence that Meese and Rogoff (1983) is still considered as a consensus in modeling foreign exchange rates for predictions. As we mentioned above, these errors and anomalies make the data more difficult to find latent patterns with neural networks. It is called aleatoric uncertainty and it does not vanish with more amount of data because it is innate uncertainty of the data itself. We treat this uncertainty with VAE and this algorithm helps training neural network and also detecting market anomalies as well.

VAE smoothed day-on-day changes Korean won/US dollar exchange rates are more stationary compared to the original series. As in Figure 8, the plot of day-on-day changes show extreme values are filtered out of the original series, and the histograms display more narrowed support of the distribution. Once VAE algorithm returns 100 filtered series, we use the mean of the series to train a deep neural network. Again with test sets of 100 filtered series, we can have 100 different predictions of Korean won/US dollar exchange rates. We therefore have the distributions and anomalies detected by the predictions.

We use Korean won/US dollar daily exchange rates as well as nine daily exchange rates either from advanced (Euro British pound, Australian dollar, Swiss franc, Japanese yen, Canadian dollar) and emerging economies (Mexican peso, Brazilian real, Indian rupee) from January 2000 to August 2019. Due to Korean foreign exchange crisis in 1997 and changes in foreign exchange regime from managed- to free-floating, it is better for the starting point to be after 2000.

4.3. Prediction Results

The prediction results show 1) deep learning is applicable to existing, not a fancy and big, macroeconomic and financial market data, 2) we can extract more information such as asymmetric risks in foreign exchange market through deep learning approach, 3) when it is compared to conventional econometric approaches, deep learning shows more prediction powers.

4.3.1. Monthly Korean Customs Clearance Exports with Ensemble Networks

The one-month ahead predictions for Korean customs clearance exports with ensemble learning approach show reasonable outcomes as in Figure 7. The error bands widen and tighten along the way, however, compared to VAR results, the error bands are much narrower which should be more informative than not. About 80%[24] of realized values are within two standard deviation of ensemble ranges. Nevertheless, the realized export values seem to follow

[24]Due to many complex features combined with Korean customs clearance exports such as foreign exchange rates or monthly business days, the realized values often deviate from the prediction range.

the prediction range very well and the predicted values point out the pivotal months of the realized values quite well. Still it is hard to conclude that the width of the range represents latent risks concerned with Korean exports as we use ensemble learning to control innate epistemic uncertainty of the series.

4.3.2. Daily Korean Won/US Dollar with VAE

The results of deep learning with VAE filtered daily Korean won/US dollar exchange series are presented as Figure 9. Compared to the prediction result from VECM with rolling windows regression with fixed starting point at the beginning of January 2000, deep neural network predictions with VAE filter show tighter error bands and more accurate one-day ahead prediction. The MAE and RMSE for VECM with rolling regression are 4.95 and 6.82 respectively whereas for deep learning, those are 3.80 and 5.26 respectively. More interestingly, the error bands out of deep learning seem to provide asymmetric risk measures for the Korean won/US dollar foreign exchange market. During the periods of time, dramatic expansions of the error bands coincide with those periods when the financial and foreign exchange markets have confronted unique events. In Figure 9, "A" period represents 7th and 8th of July 2018 when a Canadian authority arrested CFO of Huawei upon a request by the US government. At that time, "risk-off" sentiment was prevail in the financial markets due to low expectation of the US-China trade dispute reconciliation. "B" period is during 13th and 14th of May 2019 when the market expected Chinese yuan will be depreciated over 7 against US dollar. The foreign exchange market became more volatile and Korean won exchange rates oscillated in 1,200 to 1,220 range during August 2019 ("C" period) due to prospect calling off US-China trade deal, Hong Kong conflict and Chinese yuan depreciation over 7 against US dollar. Therein we can witness that the error bands do not merely display symmetric statistical confidence levels as in conventional econometric approaches. This is due to nonlinear architecture in deep neural networks which enables detecting asymmetric measure of anomaly in the time series.

5. Concluding Remarks

We present examples of deep learning applications to one-step forward predictions of macroeconomic and financial market variables. Macroeconomic time series are updated once a month or a quarter normally, and that will place some restriction applying deep learning approaches to macroeconomic data. For financial market variables, even though the frequency is high enough to be used in deep learning, they contain errors and noises within the series.

More elaborated applications, for example ensemble learning and VAE, therefore needed for filtering and processing such data.

Compared to conventional econometric approaches such as VECM, deep learning approach shows not only better prediction powers but also more informative error bands which may contain periodical development of uncertainty in the economy. In this regard, we can conclude the usefulness of deep learning even with those data which seems not to be coordinated by machine learning approach.

While this study suggests interesting future research avenues, there are some points still need to be improved. With deep learning approach, we are capable of dealing with complex big data set, so the algorithm employed in this study also can be improved by enhanced data. This study is no more than prototypes with rather arbitrary data which seems to be related with the target data. Other than the examples presented by this paper, there are a lot of data we can exploit if we adopt deep learning broadly. Most probably, we also can detect anomalies in financial markets in real time basis, examine feedback of market reaction to policy announcement and so on, if we combine existing time series with more variety of data sources such as text. We hope that this study, as a starting point, demonstrates that deep learning approach can be an alternative toolbox in addition to conventional econometrics to analyze and predict macroeconomic variables and many others.

References

Athey, S. and G. Imbens (2019): "Machine Learning Methods Economists Should Know About," *arXiv:1903.10075*.

Chakraborty, C. and A. Joseph (2017): "Machine Learning at Central Banks," *Bank of England Staff Working Paper*, 674.

Cox, D. R. (1958): "The Regression Analysis of Binary Sequences (with discussion)," *Roy Stat Soc B.*, 20, 215–242.

Doersch, C. (2016): "Tutorial on Variational Autoencoders," .

Friston, K., J. Mattout, N. Trujillo-Barreto, J. Ashburner, and W. Penny (2006): "Variational Free Energy and the Laplace Approximation," *NeuroImage*, 34.

Hinton, G. E., S. Osindero, and Y.-W. Teh (2006): "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, 18, 1527–1554.

Ho, T.-W. (2019): "Machine Learning is Not as Good as You Think for Time Series Forecasting: Evidence from Multi-Step Forecasting," *SSRN*.

Lopez de Prado, M. (2018a): "The 10 Reasons Most Machine Learning Funds Fail," *mimeo*.

——— (2018b): "Advances in Financial Machine Learning," in *Advances in Financial Machine Learning*, New York, NY, USA: Wiley.

McCulloch, W. and W. Pitts (1943): "A Logical Calculus of Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 5, 155–133.

Meese, R. and K. Rogoff (1983): "Empirical Exchange Rate models of the Seventies: Do They Fit out of Sample?" *Journal of International Economics*, 14, 3–24.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986): "Learning Representations by Back-propagating Errors," *Nature*, 323, 533–536.

Table 1: Summary statistics of year-on-year changes

Series	Mean	Standard deviation	mean (scaled)	Standard deviation (scaled)
Korean nominal export	10.529	15.093	0.511	0.171
University of Michigan consumer sentiment index	1.860	12.508	0.439	0.154
Economic policy uncertainty index (US)	8.448	44.353	0.217	0.132

This table shows summary statistics of year-on-year changes in each time series used to train and validate the ensemble of neural networks for Korean nominal export prediction. The scaled features represent the summary statistics of transformed series through mini-max scaler which standardize the year-on-year changes within [-1, 1] range.

Table 2: Summary statistics of day-on-day changes

Series	Mean	Standard deviation	mean (VAE smoothed)	Standard deviation (VAE smoothed)
Euro	0.004	0.612	-0.182	0.094
British pound	-0.004	0.594	0.203	0.055
Australian dollar	0.003	0.802	-0.006	0.061
Swiss franc	-0.003	0.142	0.056	0.033
Japanese yen	0.002	0.629	0.187	0.056
Canadian dollar	-0.000	0.564	0.125	0.075
Mexican peso	0.018	0.676	-0.164	0.055
Brazilian real	0.022	1.023	0.009	0.052
Indian rupee	0.011	0.446	-0.073	0.037
Korean won	0.004	0.681	0.126	0.048

This table shows summary statistics of day-on-day changes in each time series used to train and validate the variational encoder to estimate the joint distribution of the latent vectors. The extracted the latent vectors are used, again, to train and validate the multi-layer perceptron model for Korean won/US dollar exchange rate. The smoothed series represent the summary statistics of the series filterated by the variational autoencoder.

Table 3: Stationary Tests

Series	Test statistic	p-value
Korean exports predictions		
Korean custom clearance export	2.809***	0.999
University of Michigan consumer sentiment index	-2.930	0.042
Economic policy uncertainty index (US)	-6.334	0.000
Korean won/US dollar daily exchange rate predictions		
Euro	-1.708	0.427
British pound	-1.648	0.458
Australian dollar	-1.390	0.587
Swiss franc	-0.302	0.925
Japanese yen	-1.293	0.632
Canadian dollar	-1.346	0.608
Mexican peso	-0.465	0.899
Brazilian real	-0.391*	0.912
Indian rupee	0.155**	0.970
Korean won	-2.286	0.177

This table shows stationary test results by augmented Dickey-Fuller test with sample size = 463 for the prediction of Korean custom export and 4926 for Korean won/US dollar daily exchange rates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1.	Codependence (Presentation Slides) Marcos Lopez de Prado Cornell University - Operations Research & Industrial Engineering	1,450
2.	Clustering (Presentation Slides) Marcos Lopez de Prado Cornell University - Operations Research & Industrial Engineering	1,233
3.	Machine Learning Treasury Yields Zura Kakushadze and Willie Yu Quantigic Solutions LLC and Duke-NUS Medical School - Centre for Computational Biology	332
4.	From Generalized Linear Models to Neural Networks, and Back Mario V. Wuthrich RiskLab, ETH Zurich	220
5.	Option Pricing With Machine Learning Daniel Alexandre Bloch Université Paris VI Pierre et Marie Curie	160
6.	An Evaluation of Alternative Multiple Testing Methods for Finance Applications Campbell R. Harvey, Yan Liu and Alessio Saretto Duke University - Fuqua School of Business, Purdue University and University of Texas at Dallas - School of Management - Department of Finance & Managerial Economics	152
7.	Text Selection Bryan T. Kelly, Asaf Manela and Alan Moreira Yale SOM, Washington University in St. Louis - John M. Olin Business School and University of Rochester - Simon Business School	113
8.	Machine Learning is Not as Good as You Think for Time Series Forecasting: Evidence from Multi-Step Forecasting Tsung-Wu Ho National Taiwan Normal University, College of Management	100
9.	경제 분석을 위한 텍스트 마이닝 (Text Mining for Economic Analysis) Soohyon Kim, Young Joon Lee, Jinyoung Shin and Ki Young Park The Bank of Korea, Stanford University - Precourt Institute for Energy, Yonsei University and Yonsei University	92
10.	Neural Network for Pricing and Universal Static Hedging of Contingent Claims Vikranth Lokeshwar, Vikram Bhardawaj and Shashi Jain *affiliation not provided to SSRN*, *affiliation not provided to SSRN* and Indian Institute of Science (IISc) - Deptartment of Management Studies	86

Fig. 1. Top Downloaded Paper in Econometric and Statistical Methods on SSRN (Social Science Research Network) for the last 60 days as of January 28th, 2020

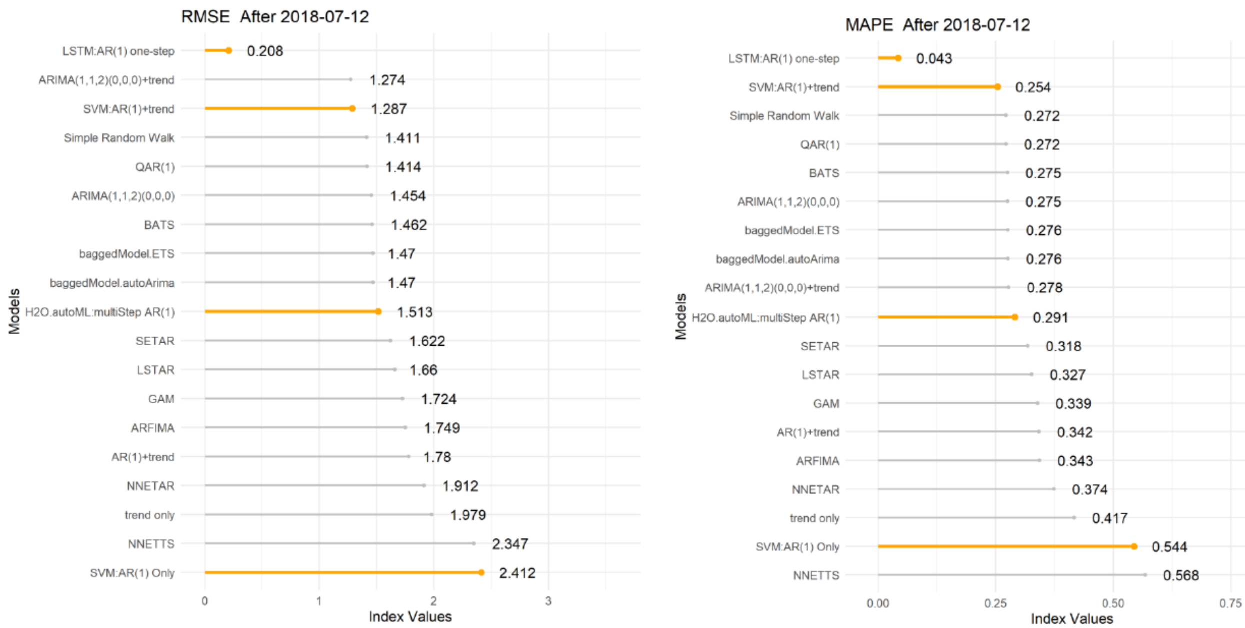


Fig. 2. Performance comparison among various approaches in econometrics and machine learning (Ho, 2019)

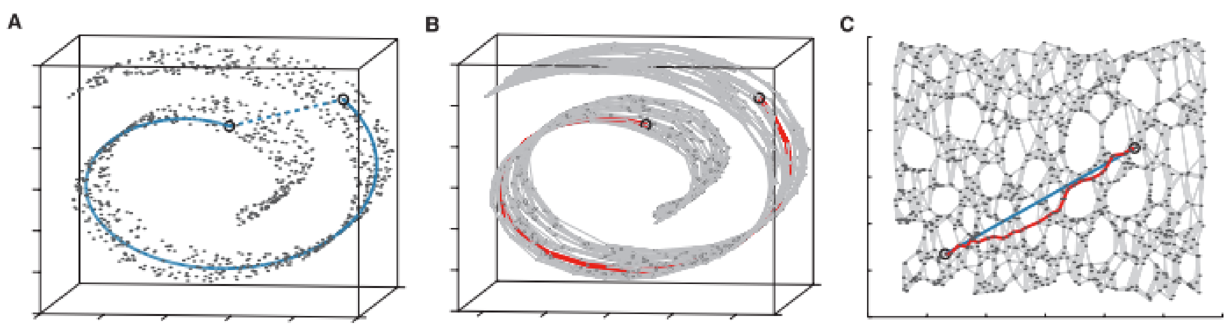
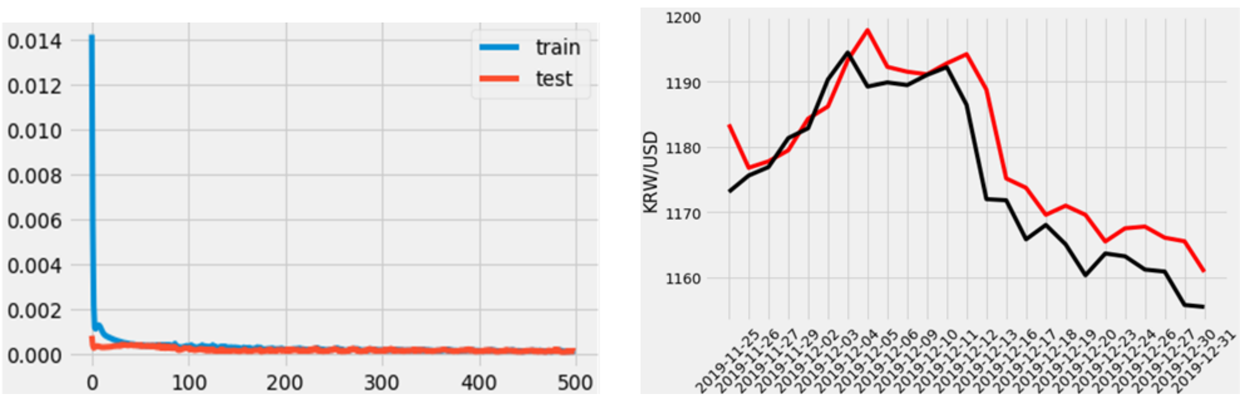


Fig. 3. Manifold Hypothesis



(a) Learning curve

(b) Over-fitted outcomes

Fig. 4. Example of over-fitting in deep learning predictions

Panel (a) shows a learning curve we can get from any training and validation procedure. Panel (b) shows one of the outcomes from what is actually over-fitted neural networks when characteristics of series are not properly addressed by neural network approach.

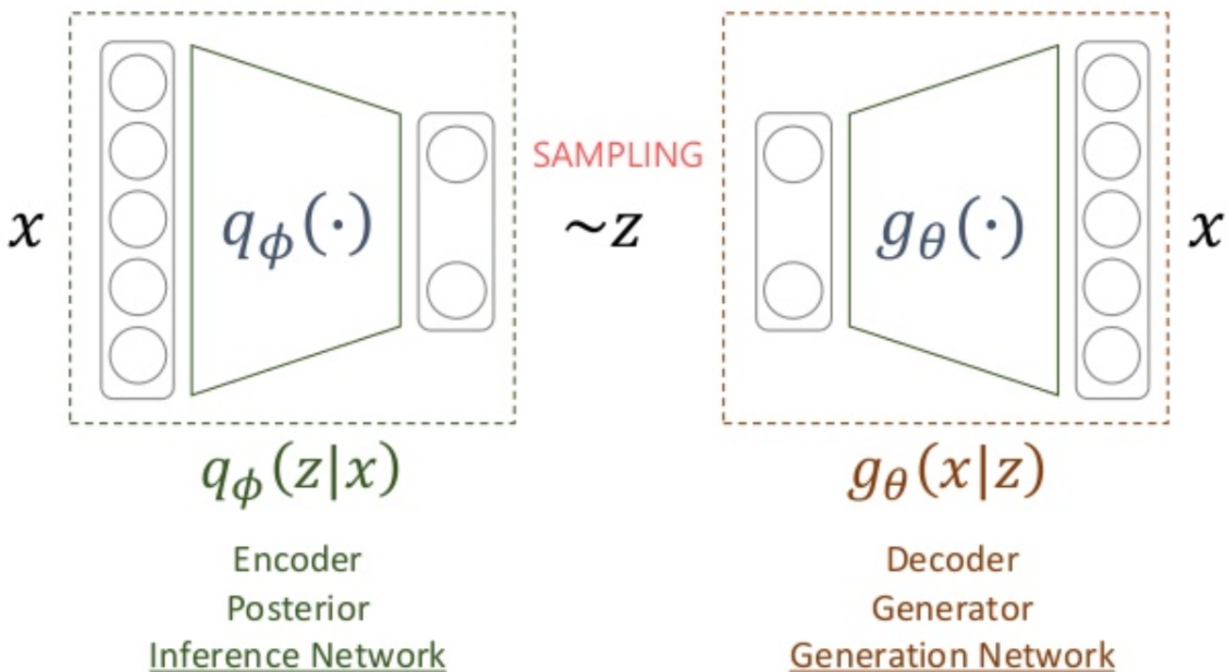
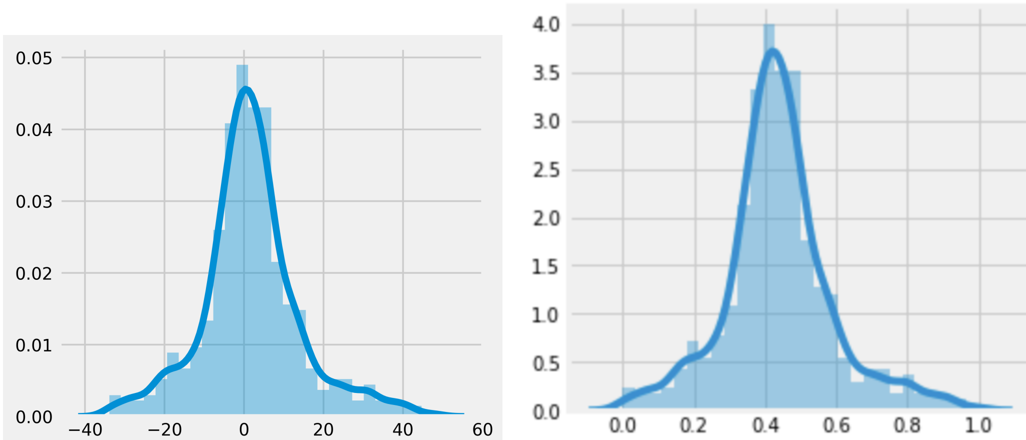
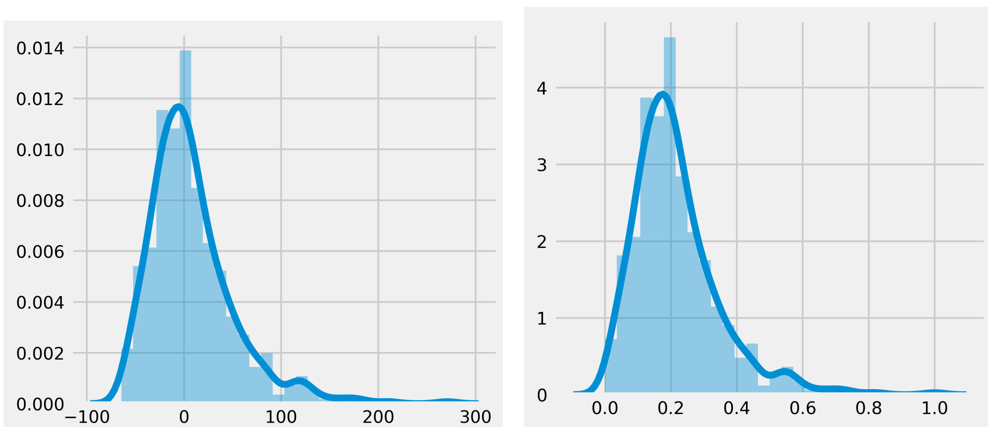


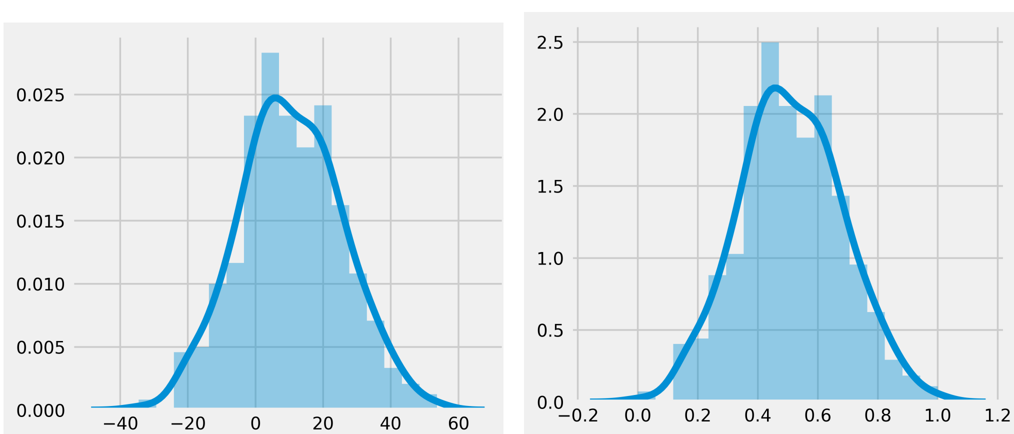
Fig. 5. Structure of Variational Autoencoders



(a) Percentage Changes of U. Michigan Index



(b) Percentage Changes of EPU Index



(c) Percentage Changes of Korea Nominal Exports

Fig. 6. Rescaled time series

Panels (a)–(c) show the histograms of year-on-year percentage changes of University of Michigan Consumer Sentiment Index, Economic Policy Uncertainty Index, and Korean custom clearance exports respectively. Right hand side of the panels are histograms for the original series whereas there are histograms for the re-scaled ([0, 1]) series on the left hand side.

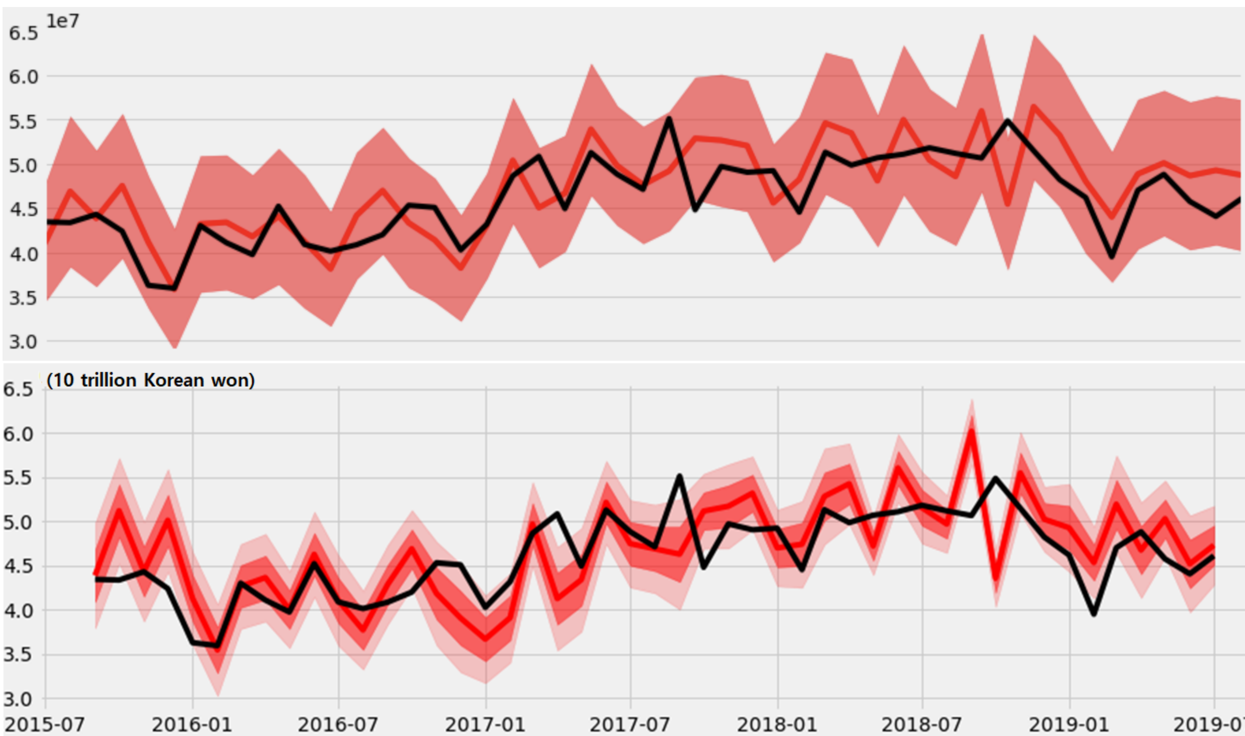
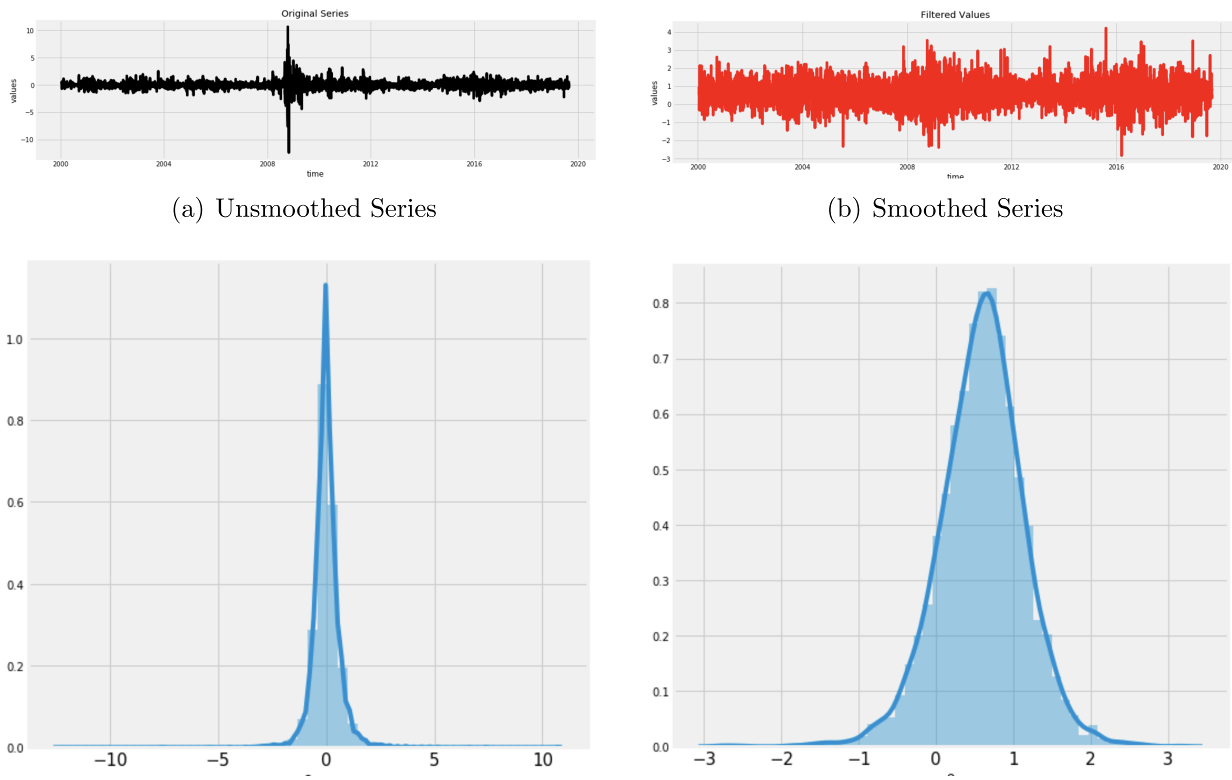


Fig. 7. Prediction of Korean custom clearance exports
Upper panel shows one-month ahead predictions of Korean custom clearance with VAR. Lower panel is the prediction results from ensemble learning. More details are on the Section 4.



(a) Unsmoothed Series

(b) Smoothed Series

(c) Histogram of unsmoothed Series

(d) Histogram of smoothed Series

Fig. 8. Smoothed Korean won-U.S. dollar exchange series with variational autoencoder
Panels (a)–(d) show the smoothed series of Korean won-U.S. dollar exchange series filtered variational autoencoder. time series of the MP sentiments of newspapers before and after MPB meetings, $tone^{before}$ and $tone^{after}$; the sentiments of MPB minutes, $tone^{minutes}$; our measure of MP surprise, $\Delta tone^{news}$; and changes in the BOK base rate, Δbase rate. Panel (e) shows the correlation coefficients.

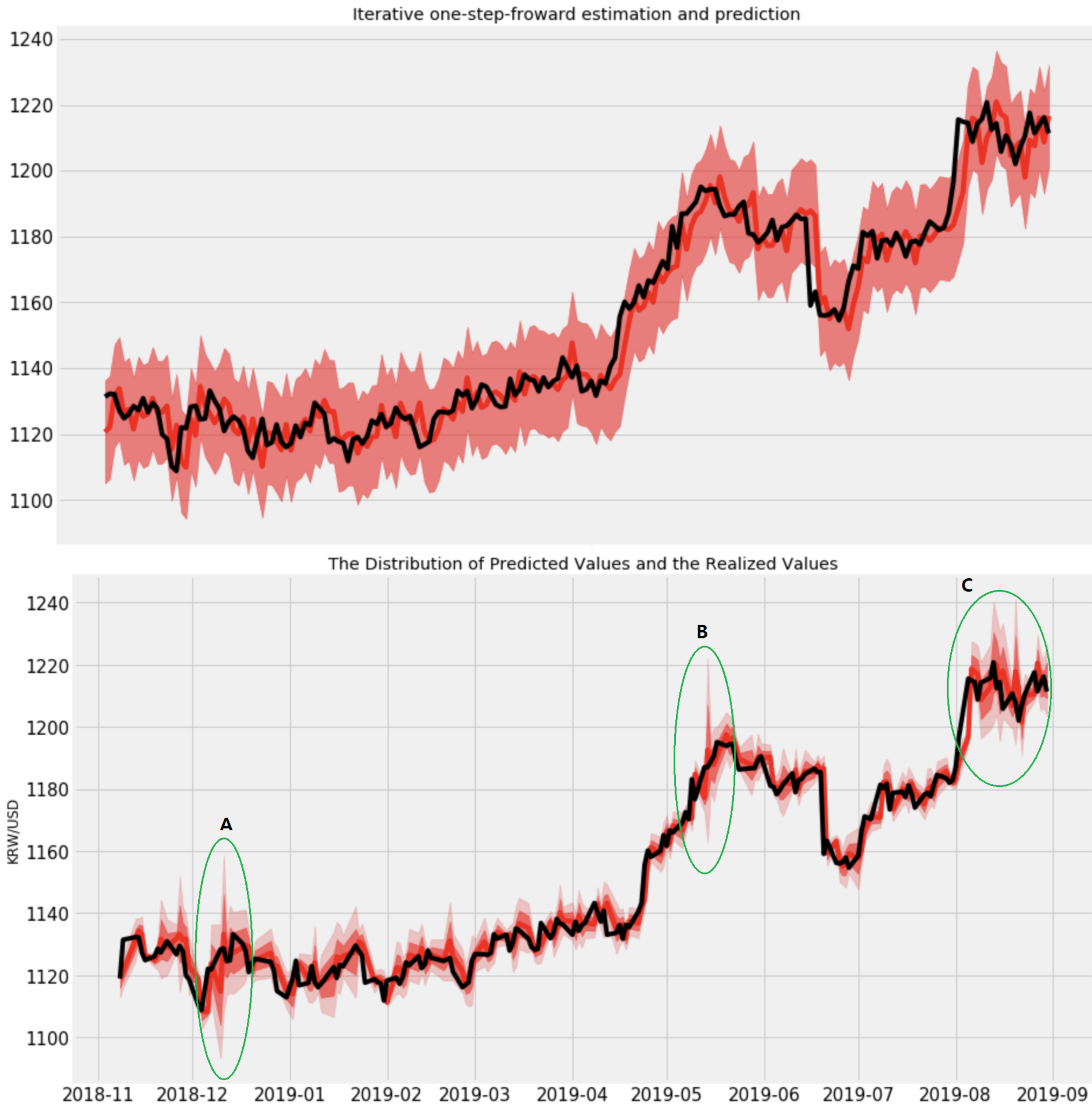


Fig. 9. Prediction of Korean won-US dollar exchange rates
Upper panel shows one-day ahead predictions of Korean won/US dollar exchange rates with VECM which is estimated based on rolling windows with fixed starting point at the beginning of January 2000. Lower panel is the prediction results from variational autoencoder (VAE) and multilayer perceptron (MLP). A, B, and C points are the periods with each unique event that has affected foreign exchange markets. More details are on the Section 4.