**Combining online search data with POS data to achieve more accurate demand forecasting-A state-space and LSTM-based approach**

Ryuji Yamamoto1, Olivia Quek Mizumura1, Wataru Ihara2

1 Data Science Team, Miotsukushi Analytics Co. Ltd., Tokyo, Japan

2 CEO and Representative Director, Miotsukushi Analytics Co. Ltd., Tokyo, Japan

Corresponding Author:

Olivia Quek Mizumura2

5F Nakajima Building, 8-8 Kabutocho Nihonbashi, Chuo-ku, Tokyo, 103-0026, Japan

Email address: olivia.mizumura@mioana.com

**Abstract**

The accurate prediction of future sales is of paramount importance to the efficient operation of any retail business since it affects the organization’s ability to plan production, procurement, inventory and promotion timings. To predict future demand in the market for cameras and vacuum cleaners in Japan, we explore the efficiency of combining online search data with Point of Sales (POS) data, to achieve one-week-ahead forecasts of consumer demand. The contribution of our paper to the existing literature is threefold. First, the existing literature is silent on whether a multivariate model, which exploits the correlation between internet search volume and product sales volume, can outperform a univariate model based on historical product sales volume only. To fill this gap in the literature, we compare the performance of univariate models with multivariate ones and exemplify how multivariate models that include internet search data outperform models that do not. Second, we attempt to not only highlight but also quantify, the effectiveness of the use of internet search data as an explanatory variable in the forecasting of future demand for retail goods. Third, we adopt relatively novel techniques, such as state-based modelling and Long Short-Term Memory (LSTM) neural networks. As far as we are aware, this is the first paper to make use of these techniques to forecast future demand for cameras and vacuum cleaners in the Japanese market.

**Introduction**

The accurate prediction of future sales is of paramount importance to the efficient operation of any retail business since it affects the organization’s ability to plan production, procurement, inventory and promotion timings. Costs can be substantially reduced with accurate forecasting, while huge losses can be incurred by poor demand estimations. While the importance of being able to accurately predict future sales cannot be over-emphasized, predictive modelling in this domain prevents various challenges. For example, the sales volumes of retail goods, such as household electronic appliances, are time-series data that typically exhibit trends and seasonal patterns. These patterns often result in volatility in sales volumes, as evidenced by highly non-stationary historical data, highly intermittent sales and irregular sales patterns.

In their attempt to forecast future prices or demand, previous studies have employed time series forecasting models, such as exponential smoothing and Autoregressive Integrated Moving Average (ARIMA)-based models. These approaches are based on univariate models that make use of previous sales data to predict future sales data. As a result, although additional time-series data, such as the volume of internet search for the products whose sales are being predicted are available, their potential ability to contribute to the forecasting of sales are ignored by these univariate models.

In recent years, several authors have attempted to apply deep learning methods to overcome challenges inherent in the forecasting of sales or demand. These state-of-the-art techniques allow them to predict demand using a multivariate approach. However, as far as we are aware, the literature is silent on whether a multivariate model, which exploits the correlation between internet search and product sales volumes, can outperform a univariate model based on product sales volumes only. This is where our paper seeks to contribute to the existing literature. In this paper, we compare the performance of univariate models with multivariate ones and exemplify how multivariate models that include internet search data outperform other models that do not.

The second contribution of our paper is its attempt to not only highlight but also quantify, the effectiveness of the use of internet search data as an explanatory variable in the forecasting of future demand. While previous works have already emphasized the benefits of using search data for predictive modelling, very few papers have compared the performance of predictive models that include search data with the performance of models that do not. Moreover, the literature is also silent on the statistical relationship between internet search data and retail sales data. In our paper, we discuss our findings on these issues within the context of the market for digital still cameras (DSCs) and vacuum cleaners (VCs) in Japan. We show that combining search data with sales data can lower the MAPE of an LSTM model by about 28% (0.036 percentage points) for the DSC market and by about 8% (0.011 percentage points) for the VC market.

Finally, the third contribution of our paper is its use of relatively novel techniques, such as state-based modelling and Long Short-Term Memory (LSTM) neural networks. As far as we are aware, this is the first paper to make use of these techniques to forecast future demand for household electronic appliances in the Japanese market.

In short, the research objective of this paper is to show that more accurate forecasting of sales or product demand can be achieved by combining internet search data with POS data. To achieve this objective, we answer the following research questions:

(1) Is search term volume a good measure of consumer demand for a retail good?

(2) How can the use of this measure reduce the error (or improve the accuracy) of a predictive model?

(3) Can highly accurate short-term forecasts of consumer demand/sales volume be achieved via relatively novel techniques, such as state-based modelling and LSTM?

**Related Works**

This paper contributes to the existing literature on online search, sales forecasting, state-space modelling and LSTM neural networks. To show how our research adds to work that has already been conducted, we briefly review the literature on information search in general, before narrowing our focus to the literature on online search. Next, we explain conventional approaches to retail sales forecasting and discuss how our methodology differs from these approaches. Finally, we review some recent works where state-space modelling and deep learning have been applied to retail sales forecasting.

**Information Search**

Information search is an area that has garnered much attention in the economics and marketing literature. For example, by creating a formal model of consumer search behavior, Meyer (1982) identifies three conditions that motivate consumers to engage in information search actively-the availability of positive information about a product, the presence of uncertainty about the product, and low search costs. Focusing on the presence of consumer uncertainty, Urbany et al. (1989) differentiate between knowledge uncertainty and choice uncertainty, where the former refers to uncertainty regarding products and the latter refers to uncertainty regarding the most optimal choice. In a similar light, Moorthy et al. (1997) identify relative brand uncertainty and individual brand uncertainty as two motivating factors that drive consumers to engage in information search. According to them, relative uncertainty has to do with which brand has the best to offer, while individual brand uncertainty has to do with what exactly each brand has to offer.

These papers differ from our research in two ways. First, they are mostly preoccupied with clarifying the underlying motivation behind consumer search behavior, rather than with predictive modelling. In contrast, our focus is on the use of search data as a measure to forecast future sales. Second, while these papers adopt a mainly theoretical approach, our approach is quantitative.

**Internet Search Data**

Moving away from the theoretical literature and in the direction of the empirical literature, previous works have shown that internet search data can be an effective marketing metric for economic data releases. For example, Wu et al. (2009) find that internet search counts can be used to predict house prices and sales in the U.S., and Askitas et al. (2009) discover strong correlations between keyword searches on the internet and the unemployment rate in Germany. Several authors have also examined the use of internet search data in forecasting private consumption. A noteworthy example would be Kholodilin et al. (2010)’s paper, which compares the performance of an internet search-based forecasting model against that of several benchmark models of private consumption that do not include search data in their list of independent variables. The authors discover statistically significant evidence in favor of their hypothesis that the internet search-based model outperforms benchmark models in terms of predictive accuracy.

While the above-mentioned works have exemplified the effectiveness of internet search data as a metric for forecasting, most of them focus on the prediction of macroeconomic indicators, such as the national unemployment rate and national private consumption, rather than on the prediction of microeconomic indicators.

At the time of writing this paper, research on the use of internet search data for the prediction of microeconomic indicators is still in its early stages, as evidenced by the fact that very few authors have commented on the use of search data for predicting prices or demand at the microeconomic level. Perhaps the closest works that we could find on this subject are papers where search data was used to predict stock market prices. In this field of research, authors are inconclusive on whether online search data can be a useful predictor of stock prices. For example, Preis et al. (2010) examine weekly search volume for several search terms submitted to the search engine Google between 2004 and 2010 and discover that weekly transaction volumes of S&P companies are strongly correlated with weekly search volumes of the names of these companies. However, Vlastakis et al. (2012) argue that the positive correlation between information demand and supply “does not allow conclusive inferences about the information discovery process”. Using search data as an indicator for information demand on 30 of the largest stocks traded on the NYSE and NASDAQ, they exemplify how an increase in information demand leads not only to an increase in trading volumes but also to greater volatility in the stock market. Their findings imply that the increase in volatility caused by an increase in information demand makes search data an unreliable predictor of stock prices.

Our paper distinguishes itself from the above-mentioned papers, by focusing on the forecasting of a specific microeconomic variable, that is, consumer demand for cameras and vacuum cleaners.

**Retail Sales Forecasting**

As previously mentioned, retail sales forecasting extends from a tradition of univariate forecasting methods, such as exponential smoothing and ARIMA models. Exponential smoothing is based on a description of trend and seasonality in the data, while ARIMA models seek to discover autocorrelations in the data. A shortcoming of these approaches is that although additional time-series data, such as the volume of internet search for the products whose sales are being predicted are available, their potential ability to contribute to the forecasting of sales are completely ignored.

In recent years, several authors have attempted to use state-of-the-art techniques, such as state-space modelling and deep learning, to predict demand. For example, Alon et al. (2001) used an Artificial Neural Network (ANN), a Winters exponential smoothing model, an ARIMA model and a multiple regression model to make out-of-sample forecasts of aggregated retail sales. Comparing the Mean Average Percentage Error (MAPE) of these models, they found that traditional methods, such as the Winters exponential smoothing model and the ARIMA model, outperform the ANN under relatively stable macroeconomic conditions. However, the ANN outperforms the other models when economic conditions are volatile. In the same spirit, Ramos et al. (2015) compare the performance of state-space models against ARIMA models for forecasting the retail sales volumes of five different categories of women’s footwear and discover that both methods perform quite similarly.

As highlighted by Bandara et al. (2019), a major advantage of state-of-the-art techniques such as Recurrent Neural Networks (RNN) and Long Short-Term Memory Networks (LSTM) is that they allow researchers to “exploit the enormous potentials of multiple time-related time series”. Some noteworthy attempts to forecast retail sales using a multivariate approach include Bandara et al. (2019)’s research, which exploits cross-product demand patterns and correlations to forecast the sales demand for individual products sold on *Walmart.com*; Borovykh et al. (2017)’s research, which develops a deep Convolutional Neural Network based on the WaveNet architecture and tests its performance on multivariate time series data including commodities data; and Salinas (2020)’s research, which uses *DeepAR*-a methodology based on training an Autoregressive Recurrent Neural Network model on a large number of related time series-to achieve accurate probabilistic forecasts of five real-world datasets, including weekly item sales from Amazon.

The above-mentioned works highlight the potential for multivariate models to increase the accuracy of demand forecasting. However, they do not explore the potential benefits of using internet search data as a predictor variable for consumer demand. In the sections that follow, we fill the gap in the literature by investigating whether search term volume can be a good measure of consumer demand, and the extent to which this measure can improve the accuracy of a predictive model.

**Materials & Methods**

Can a multivariate predictive model, which exploits the correlation between internet search volume and product sales volume, perform better than a univariate predictive model based on historical product sales volume only? To answer this question, we use internet search data and Point-of-Sales (POS) data provided by GFK Japan to forecast future sales in two different markets in Japan-the market for Digital Still Cameras (DSCs) and the market for vacuum cleaners (VCs). GFK is Germany’s largest market research institute and the fourth largest market research organization in the world as at the time of writing[[1]](#footnote-1).

**Data for the Digital Still Camera (DSC) and Vacuum Cleaner (VC) market**

POS data is comprised of the weekly traditional and internet sales volumes of 59 brands of DSCs and 168 brands of VCs, between the 1st week of 2017 and the 48th week of 2019[[2]](#footnote-2). There are two different measures of sales volumes-the number of units of a particular brand sold within a specific week and the total pre-tax sales volume of a particular brand sold within a specific week (measured in Japanese Yen). We make use of the latter measure for this study. To perform predictive modelling at the aggregate market level, we sum weekly traditional and internet sales volumes across all individual brands.

Internet search data provided by GFK is comprised of daily search volumes that took place on the internet search engine Yahoo! Search, between 1st January 2017 and 30th November 2019. The data for each of the respective markets were computed by aggregating the number of unique users who entered a search term that included ‘camera’ or ‘vacuum cleaner’ within each day of the given timeframe. An example of such a search term would be ‘camera/vacuum cleaner’, followed by a specific brand. Data is available at the national aggregate level as well as at a regional level[[3]](#footnote-3). In addition, details on the gender, age group and prefectural location of each unique user are also available. To standardize the frequency of the 2 time-series data, we aggregate daily search volumes to the weekly level and use weekly search and sales volumes to make one-week-ahead forecasts of DSC and VC sales.

To better understand the search and sales patterns for DSCs and VCs across time, we use ETS decomposition to split the time series data for each respective market into several components, namely, observed values, trend, seasonality and a residual component. Figure 1 and Figure 2 depict the results of ETS decomposition for DSC sales and search volumes, respectively. From these figures, we observe that both the search and sales volumes for DSCs exhibit downward-sloping trends over time and follow a quarterly seasonal pattern. Figure 3 and Figure 4 depict the results of ETS decomposition for VC sales and search volumes, respectively. The figures show that like the market for DSCs, search and sales volumes for VCs also exhibit a clear quarterly seasonal pattern.

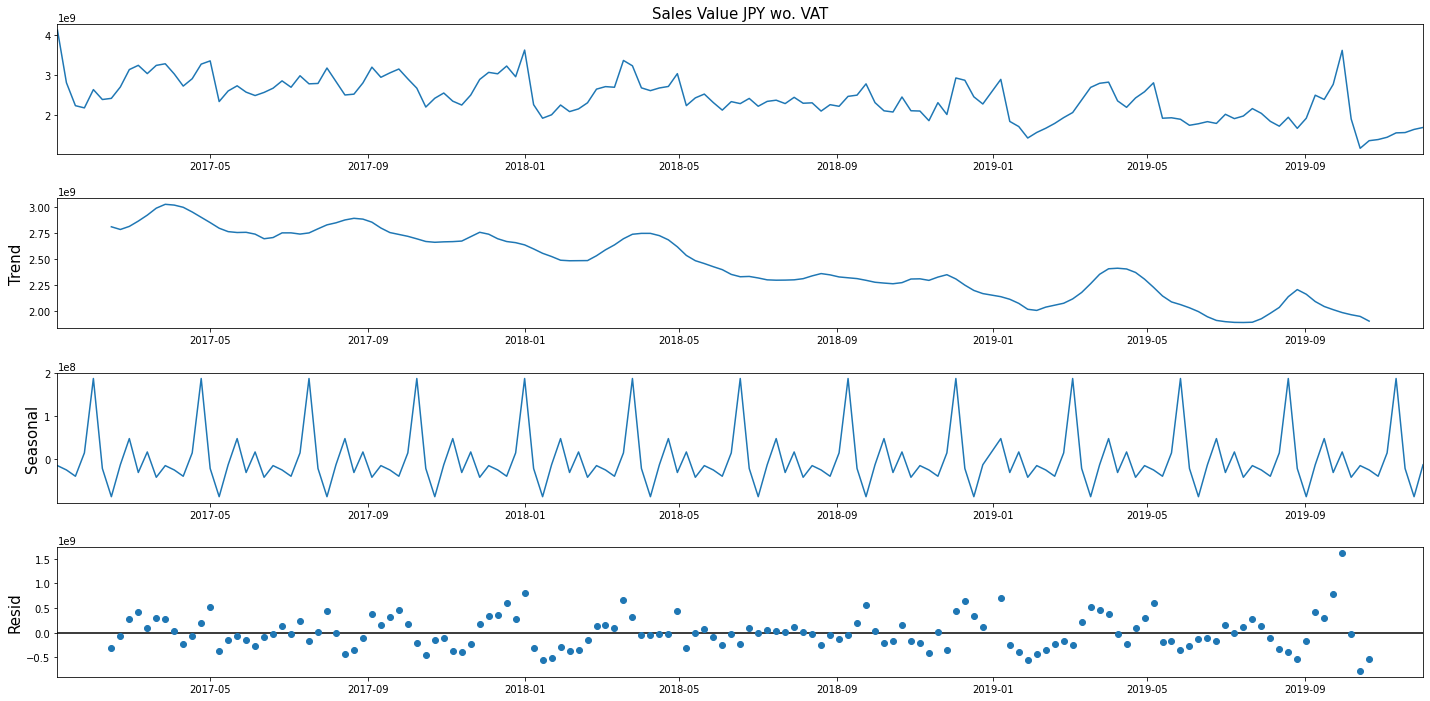


Figure 1 Observed total **pre-tax sales amount of all DSCs** sold within a specific week (in JPY), its trend, seasonality, and residual component between the 1st week of 2017 and the 48th week of 2019.

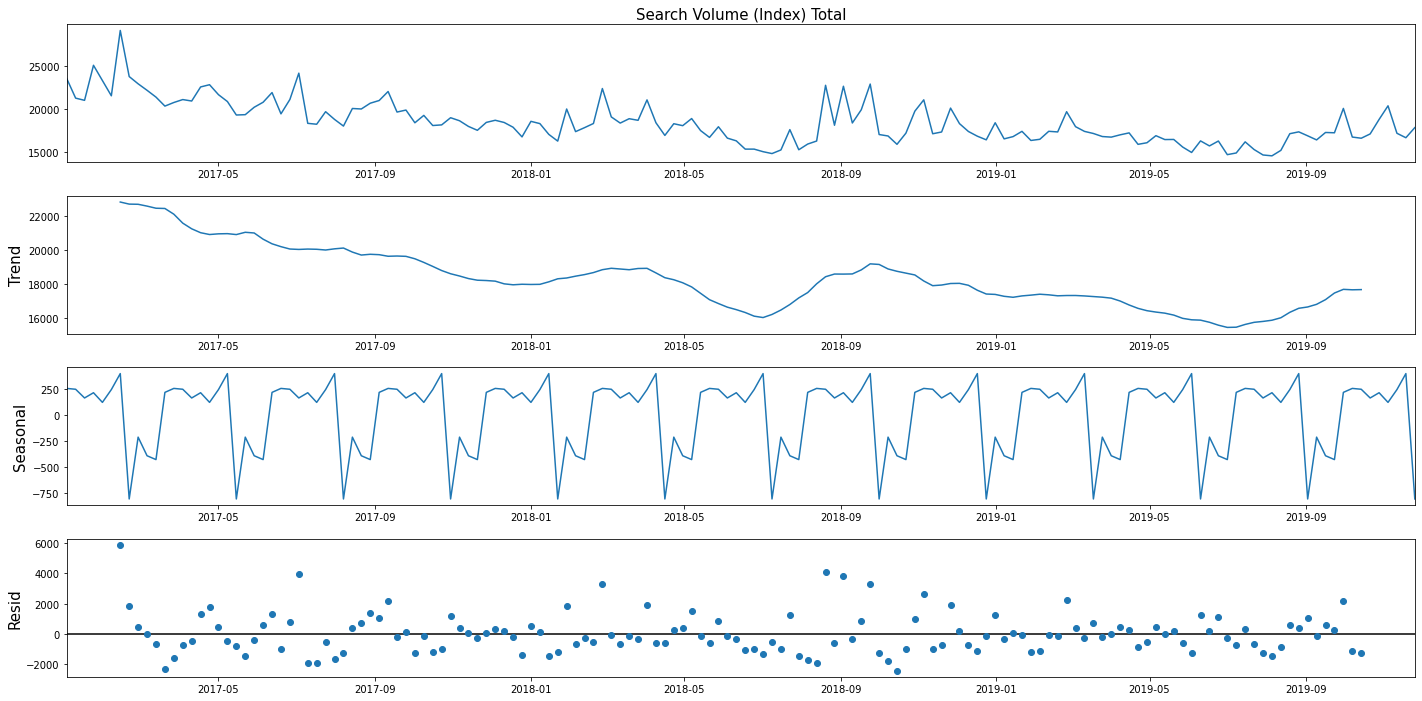


Figure 2 Observed weekly **search volume for the DSC market**, its trend, seasonality, and residual component between the 1st week of 2017 and the 48th week of 2019.

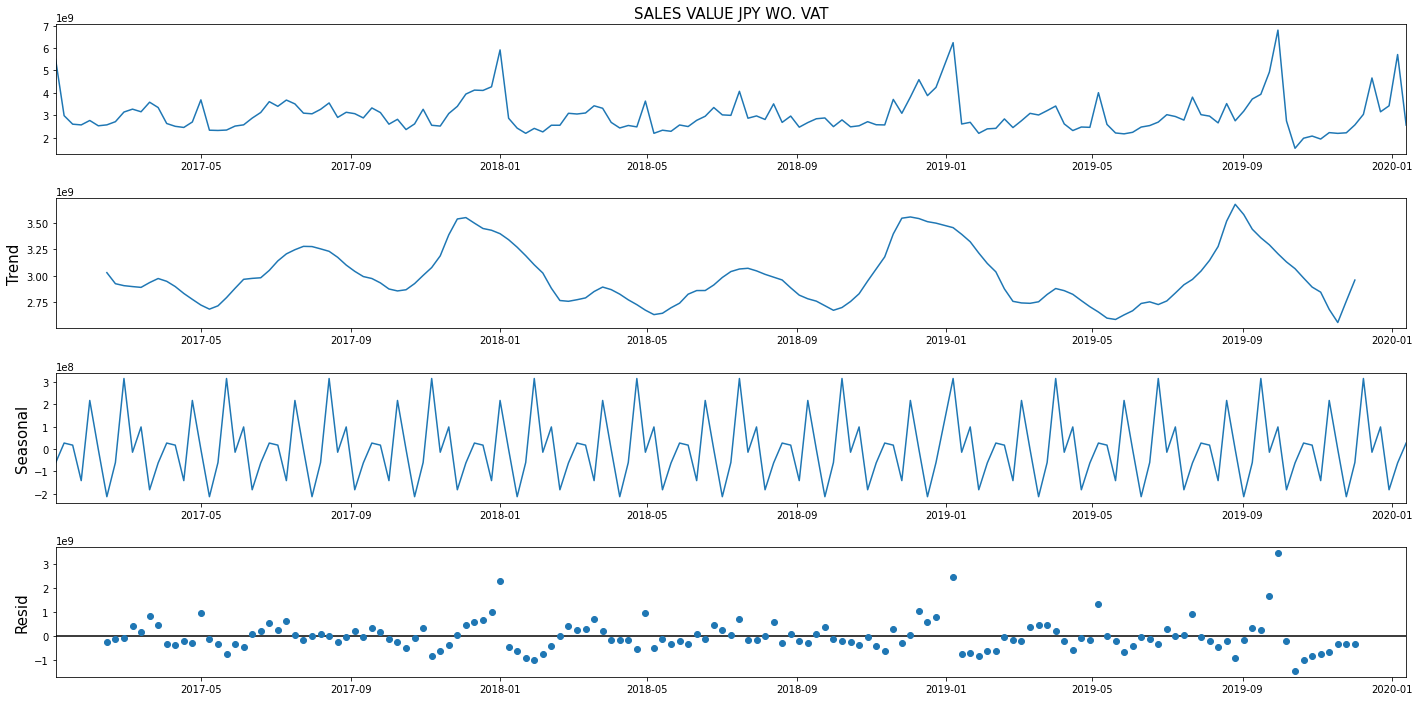


Figure 3 Observed total **pre-tax sales amount of all VCs** sold within a specific week (in JPY), its trend, seasonality, and residual component between the 1st week of 2017 and the 48th week of 2019.

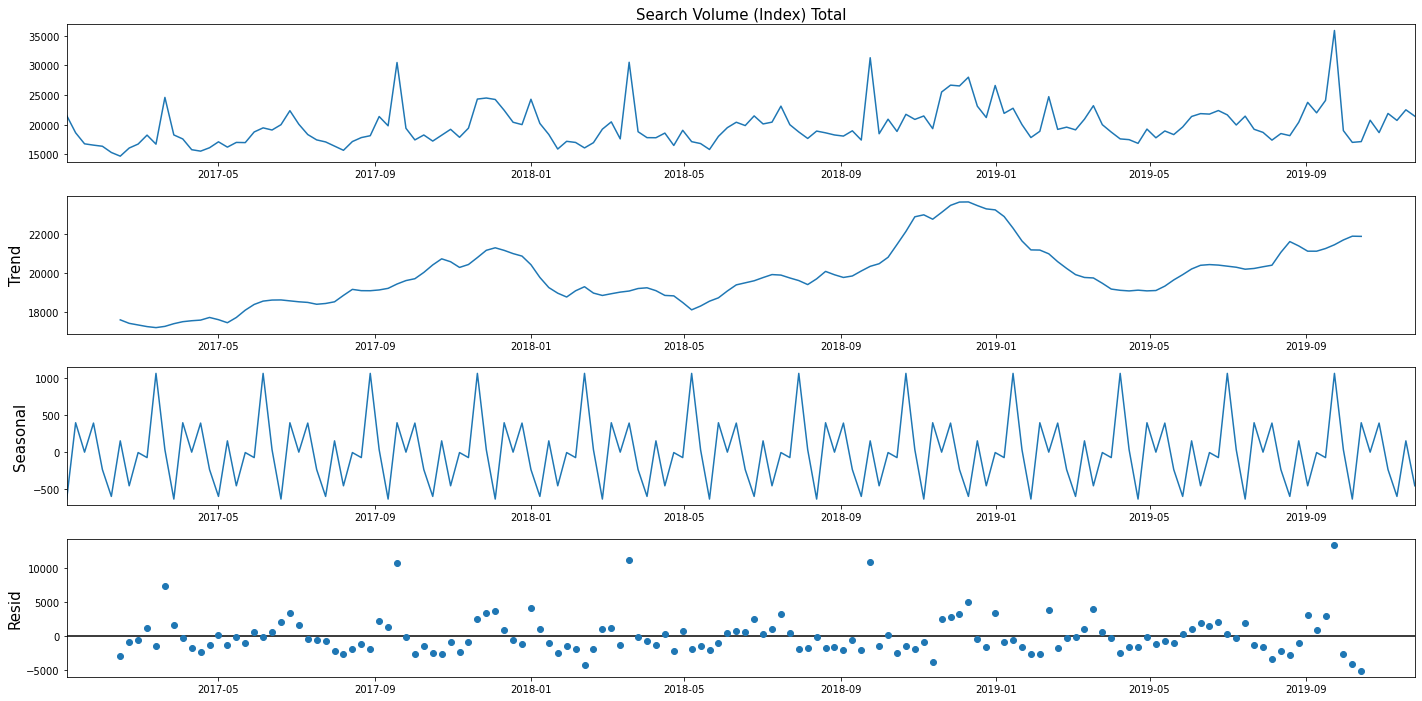


Figure 4 Observed weekly **search volume for the VC market**, its trend, seasonality, and residual component between the 1st week of 2017 and the 48th week of 2019.

**Granger causality test and Johansen test results**

Does a statistically significant relationship between search and sales data exist? To be more precise, is search data useful in forecasting sales data and vice versa? Do these time series data share a cointegrating relationship? To answer these questions, we perform Granger causality and Johansen tests on search and sales data for the DSC and VC markets, respectively.

The Granger causality test is a statistical hypothesis test that helps us discern if a time series data is useful in forecasting another. The null hypothesis for the test is that lagged search volumes do not provide statistically significant information about the evolution of future sales volumes and vice versa.

We present the results of the Granger causality test for search and sales data in the DSC market, in Table 1. Each cell in the table contains the p-value for testing the null hypothesis that a given variable denoted by the variable name followed by the suffix ‘\_X’ does not Granger-cause another variable denoted by the variable name followed by the suffix ‘\_Y’. The table shows that both the null hypotheses that (1) search volume does not Granger-cause sales volume, and that (2) sales volume does not Granger-cause search volume, can be rejected at a 5% level of significance. This is because the p-values for each of these hypotheses are 0.0022 and 0.0029 respectively.

Table 2 captures the results of the Granger causality test for search and sales data in the VC market and can be read in the same way as Table 1. The table shows that the null hypothesis that (1) search volume does not Granger-cause sales volume can be rejected at the 5% level of significance, although the null hypothesis that (2) sales volume does not Granger-cause search volume cannot. However, the fact that Granger causality was detected between search and sales data suggests that some sort of statistical relationship does indeed exist between these 2 variables, in the market for VCs.



Table 1 Granger causality test results for the DSC market. Each cell in the table contains the p-value for testing the null hypothesis that a given variable, denoted by the variable’s name followed by the suffix ‘\_X’, does not Granger-cause another variable, denoted by the variable’s name followed by the suffix ‘\_Y’.

|  |  |  |
| --- | --- | --- |
|  | **Sales Value JPY wo. VAT\_X** | **Search Volume (Index) Total\_X** |
| **Sales Value JPY wo. VAT\_Y** | 1.0000 | 0.0000 |
| **Search Volume (Index) Total\_Y** | 0.1748 | 1.0000 |

Table 2 Granger causality test results for the VC market. Each cell in the table contains the p-value for testing the null hypothesis that a given variable, denoted by the variable’s name followed by the suffix ‘\_X’, does not Granger-cause another variable, denoted by the variable’s name followed by the suffix ‘\_Y’.

A stationary time series is one whose properties are non-dependent on the time at which the data is observed. This means that time series with trends or seasonality are not stationary, since the trend and seasonality will affect the value of the data depending on the time it is observed. The Johansen test helps us determine if cointegration is present between 2 non-stationary time series data. Cointegration is present when 2 or more nonstationary time series data have a long-run equilibrium, move together in a way that their linear combination results in a stationary time series, and share an underlying common stochastic trend.

We perform the Johansen test on search and sales data of the DSC and VC markets respectively, to test the null hypothesis that no cointegrating relationship exists between search and sales data. Table 3 shows the results of the test for the DSC market. Since the test statistic of 21.92 exceeds the critical value of 12.32 at the 95% confidence level, we can reject the null hypothesis and conclude that cointegration is present between search and sales data for the DSC market at the 5% level of significance. Similarly, Table 4 shows the results of the test for the VC market. Since the test statistic of 21.43 exceeds the critical value of 12.32 at the 95% confidence level, we can reject the null hypothesis and conclude that cointegration is present between search and sales data for the VC market at the 5% level of significance.

To this end, we performed the Granger’s Causality test and Johansen test on sales and search data of the markets for DSCs and VCs in Japan and were able to ascertain from the results of these tests that the 2 time-series data do share a statistically significant relationship.



Table 3 Results of the Johansen cointegration test performed on search and sales data of the DSC market. Since the test statistic of 21.92 exceeds the critical value of 12.32 at the 95% confidence level, we can reject the null hypothesis and conclude that cointegration is present between search and sales data for the VC market at the 5% level of significance.



Table 4 Results of the Johansen cointegration test performed on search and sales data of the VC market. Since the test statistic of 21.43 exceeds the critical value of 12.32 at the 95% confidence level, we can reject the null hypothesis and conclude that cointegration is present between search and sales data for the VC market at the 5% level of significance.

**Accounting for the lagged response of sales data with respects to search data**

It is logical to assume that most customers would search for information on the latest models of DSCs and VCs for quite a few weeks before purchasing them in stores or on e-commerce websites. In other words, it is highly plausible that sales data responds to search data only after one or more weeks after the search data was observed. To check if such a lag exists, we pair each week’s sales volume with the search volume that was observed one or more weeks ago. Next, we perform the Johansen test on weekly sales data and time-shifted search data for the DSC and VC markets, respectively. This allows us to observe how the Johansen test rejection rate changes with respects to the number of weeks by which search data is shifted. Figure 5 provides a graphical representation of data that accounts for a 2-week lagged response of sales data with respects to search data.

**Original data**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** | **Week 6** | **…** |
| **Search**  **volume** | 145 | 139 | 132 | 129 | 134 | 135 | … |
| **Sales**  **volume** | 413 | 280 | 222 | 217 | 263 | 238 | … |

**Data with search volume shifted by 2 weeks**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** | **Week 6** | **…** |
| **Search**  **volume** | - | - | 145 | 139 | 132 | 129 | … |
| **Sales**  **volume** | 413 | 280 | 222 | 217 | 263 | 238 | … |

Figure 5 A graphical representation of data that accounts for a 2-week lagged response of sales data with respects to search data.

Figures 6 and 7 show how the test statistic for the Johansen test changes with the number of weeks by which search data is shifted, for each respective market. The critical value remains unchanged at 12.32. For the DSC market, shifting search data with respects to sales data by 10 weeks leads to the highest rejection rate of the Johansen test’s null hypothesis. For the VC market, shifting search data with respects to sales data by 7 weeks leads to the highest rejection rate. Based on the results of this analysis, we consider a 10-week lagged response of sales data with respects to search data for the DSC market, and a 7-week lagged response of sales data with respects to search data for the VC market when building state-space and LSTM models for predictive modelling.

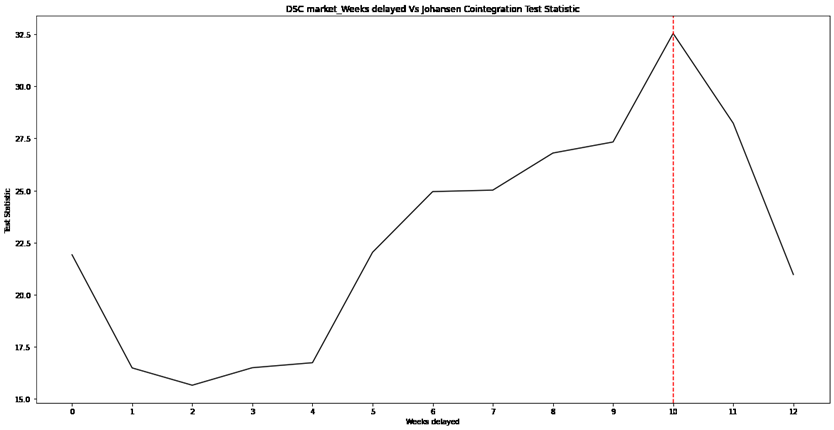


Figure 6 Plot of the number of weeks by which the DSC market’s search data is shifted (x-axis), against the Johansen test’s test statistic (y-axis).

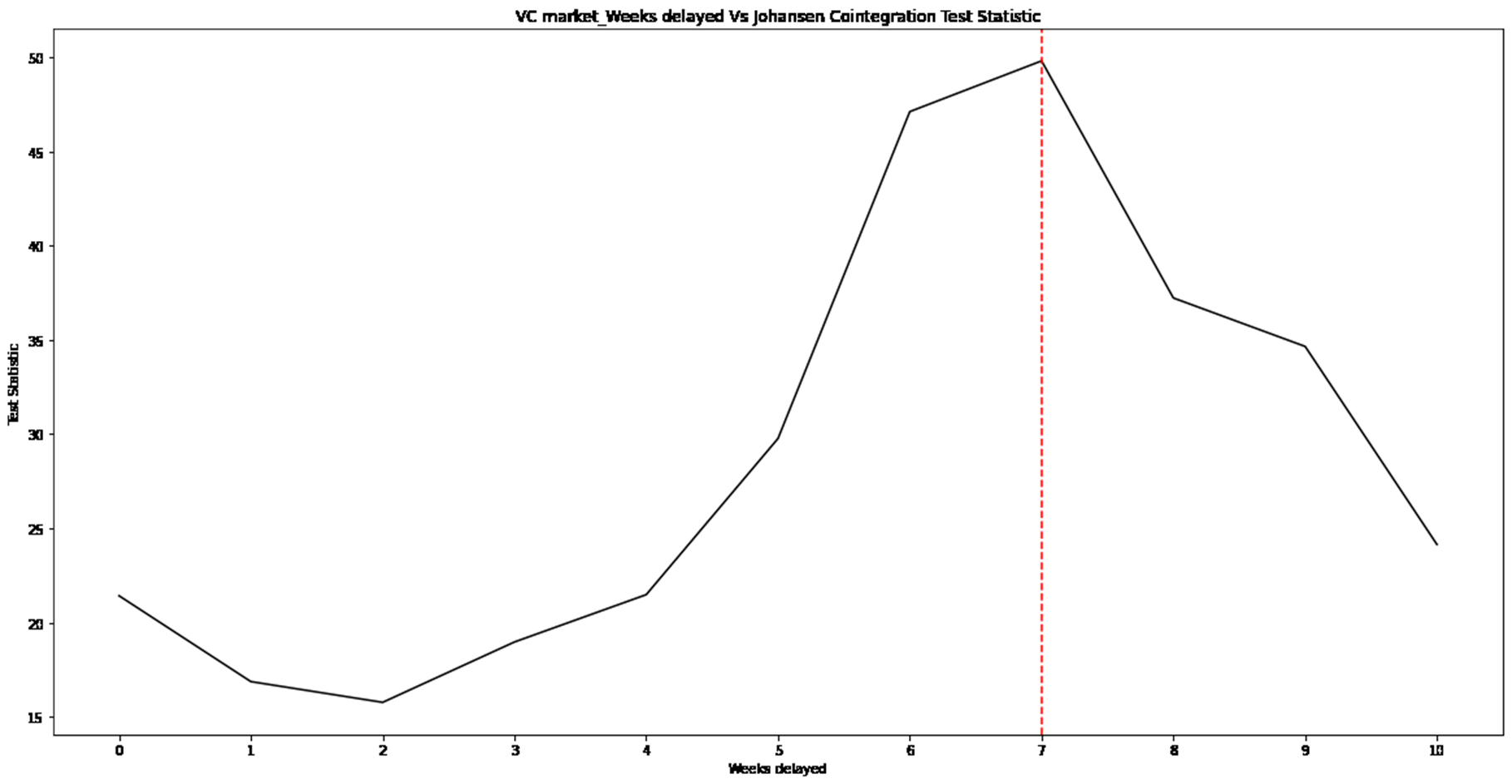


Figure 7 Plot of the number of weeks by which the VC market’s search data is shifted (x-axis), against the Johansen test’s test statistic (y-axis).

**State-space models**

In this sub-section and the next, we explain the methods we used to produce one-week-ahead forecasts of aggregate demand/sales volumes in each of the markets of our interest. To answer the research question of whether a multivariate model, which exploits the correlation between internet search volume and product sales volume, can outperform a univariate model based on historical product sales volume only, we first construct several state-space models and compare the performance of the univariate models against the multivariate ones. The results of this comparison allow us to conclude that combining search data with sales data does indeed improve the accuracy of the forecasts.

State-space models do not have a fixed form and encompass all models that describe the probabilistic dependence between one or more latent, that is, unobserved, state variable(s) and one or more observed variable(s). In mathematical notation, the general form of a state-space model can be written as follows, where the first equation captures the movement of the observed variable and the second captures the movement of the latent state variable .

At each time , the observed variable is the sum of , which is a function of the latent variable, and noise . The latent variable is the sum of , which is a function of its value in the previous time and its noise at time , .

Figure 8 graphically depicts these state-space equations. The figure shows how state-space modelling encompasses the following 2 steps of inference.

1. Prediction. This means using previous values of the state to forecast subsequent values of the state.
2. Filtering. This involves estimating the current values of the state from past and currently observed variables.

In the context of this paper, we can think of the state variable as being the aggregate market demand for DSCs and VCs, and the observed variables as being the sales and search data for DSCs and VCs at each point in time of our dataset.

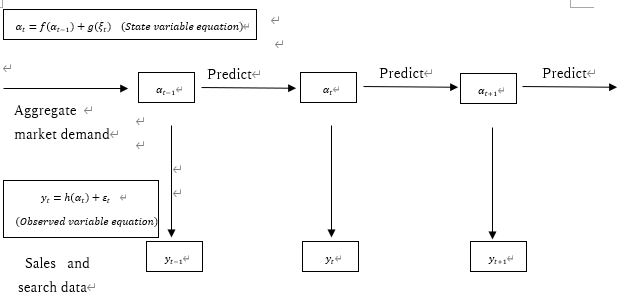


Figure 8 Graphical representation of the prediction and filtering processes of a state-space model

For each of the markets of our study, we construct 4 univariate state-space models, where historical sales volumes are used to make one-week-ahead forecasts of aggregate market demand, and 2 multivariate state-space models, where historical sales and search volumes are used to make one-week-ahead forecasts of aggregate market demand. Tables 5 and 6 summarize the types of models we constructed, and the explanatory variable(s) used in each of the models to forecast market demand. In what follows, we shall briefly describe each of these models and explain how they differ from one another.



Table 5 Summary of univariate state-space models constructed in this study.



Table 6 Summary of multivariate state-space models constructed in this study.

Local level model

The local level model belongs to a class of state-space models known as unobserved components models. Unobserved components models decompose a univariate time series into several components, namely the trend, seasonal, cyclical and idiosyncratic components while controlling for the influence of exogenous variables. The local level model assumes that the latent state follows a random walk and that the observed variable is the result of the latent state plus noise. It can be described mathematically as follows,

where is the value of the observed variable at time, is the value of the latent state variable at time , and and are noise variables associated with the observed and latent variables, respectively.

Local linear trend model

The local linear trend is another form of an unobserved components model. It decomposes the trend component of a univariate time series into 2 separate components, namely, a deterministic trend component and a probabilistic trend component. It can be described mathematically as follows,

where is the value of the observed variable at time, is the value of the latent state variable at time , is the probabilistic trend component, and , and are the noise variables associated with the observed, latent, and probabilistic trend variables, respectively.

Local level model with seasonality

This model incorporates the consideration of seasonality into the basic local level model. For quarterly data, such a seasonal component could be modelled with the following framework,

where the ’s refer to the seasonal components and the disturbance allows the seasonal component to change over time. The framework above assumes that:

(1) the observed variable depends on the state variable , the seasonal component term and a noise term ;

(2) the state variable depends on its previous value and a noise term ; and

(3) the seasonal component changes over time based on its past values and the disturbance term .

The last 2 equations represent identities and imply that follows and follows .

Local linear trend model with seasonality

Next, we incorporate seasonality into the local linear trend model that was defined above. An example of such a framework, which assumes that data is quarterly, can be denoted as follows,

Where is the value of the observed variable at time, is the value of the latent state variable at time , the ’s refer to seasonal components and is the probabilistic trend component. The seasonal component changes over time based on its past values and the disturbance term . , , and are the disturbance terms associated with the observed variable, latent variable, probabilistic trend component and seasonal component, respectively.

Univariate autoregressive models represented in state-space

Univariate autoregressive models forecast a variable of interest using a linear combination of the variable’s historical values. These models are also known as the ARIMA class of models, where ARIMA is an acronym that stands for **A**uto-**R**egressive **I**ntegrated **M**oving **A**verage. Each component of the acronym is defined as follows.

-*Autoregression* (AR): refers to a model that regresses a variable against past values of itself.

-*Integrated* (I): refers to the use of a technique known as differencing, to make a non-stationary time series stationary.

-*Moving Average* (MA): refers to a model that uses the dependency of a variable against residual errors from past forecasts of itself to predict future values of the variable.

AR and MA models work on stationary time series, so integration (I) is a preprocessing procedure that needs to be applied in the case where the time series is non-stationary.

A standard notation of autoregressive models is ARIMA(p, d, q) where the parameters in the parenthesis are substituted with integer values to indicate the specific type of model used. The parameters are defined as follows.

-p: The number of periods by which the variable of interest is lagged against itself. This is also called the lag order of the model.

-d: The number of times that the raw observations are differenced, also called the degree of differencing.

-q: The size of the moving average window.

An ARIMA model can be configured to perform the function of a simpler model, such as AR, MA, or ARMA. In this case, a value of 0 is used for the parameter(s) that are not used in the model.

Additionally, we can add the consideration of seasonality to an ARIMA model. A Seasonal ARIMA model, or SARIMA model, is formed by adding additional seasonal terms to the above-mentioned ARIMA models. A standard notation of the SARIMA model is SARIMA(p, d, q)(P, D, Q)[m] where (p, d, q) and (P, D, Q) are the lag orders, degree of differencing and size of the moving average window of the non-seasonal part and the seasonal part of the model, respectively. [m] represents the number of observations per year.

In the context of this study, we represent autoregressive models in state-space form, to forecast the unobserved state variable-aggregate market demand, using the observed variable-sales volume. To show how any autoregressive model can be represented in state-space form, we use the example provided by Rothenberg (2007).

First, consider a linear state-space model that postulates that an observed time series is a linear function of an unobserved state vector. More precisely, let be the observed variable at time and be the values assumed by a vector of state variables at time . Let and be and matrices of constants. Assume that is generated by the following process,

where the scalar and the vector represent white-noise processes with mean zero and are of the initial value . Further, and .

Next, consider the ARMA(1,1) model, which is commonly denoted as follows.

Defining , we can write where and

.

From this example, we see that the ARMA(1,1) model has a state-space representation with .

More generally, we can write an ARMA(p,q) process for the observed variable with mean-zero in state-space form as follows. Let . Then, we can write

with redundant coefficients set to 0. Defining the column vectors

, , ,

where b is , c is and d is in size.

By successive substitution, we can prove that has the state space form:

, ,

where is an m-dimensional state vector, , and

.

In the context of this study, we find the best parameters for an ARIMA (or SARIMA) model for the DSC and VC markets using a grid search, where the Akaike Information Criterion (AIC) is used as an indicator of each model’s performance. Based on the results of parameter tuning, we find that an ARIMA(1,1,1) model or SARIMA(1,1,1)(1,0,0)[4] model best represents the DSC market and that an AR(1) model best represents the VC market. We use the state-space representation of each of these models to obtain one-week-ahead aggregate market demand forecasts for each respective market.

The models described thus far have been univariate. In what follows, we describe several multivariate state-space models that we constructed. These models make use of historical sales and search volumes to make one-week-ahead forecasts of aggregate market demand.

Multivariate autoregressive models represented in state-space

Multivariate autoregressive models include the Vector Autoregression (VAR) model and the Vector Autoregression Moving Average (VARMA) model. The VAR model can be thought of as the multivariate extension of the AR model. It regresses a target variable against past values of itself and past values of one or several other explanatory variables, to make forecasts of the target variable. A VAR model has the parameter p, where each equation in the VAR(p) model contains p lags of all variables in the system. For example, a VAR(1) model can be written as follows.

The VARMA model incorporates a moving average component to the VAR model. That is, it uses the dependency of a target variable against residual errors from past forecasts of itself and past forecasts of one or several other variables to predict future values of the target variable.

A VARMA model has the parameter (p,q), where p is the fixed autoregressive order and q is the moving average order. The VARMA(1,1) model can be defined mathematically as follows.

We find the best parameters for a VAR (or VARMA) model for the DSC and VC markets using a grid search, where the Akaike Information Criterion (AIC) is used as an indicator of each model’s performance. Based on the results of parameter tuning, we find that a VAR(5) model best represents both markets, when the consideration of the moving average is excluded. When the consideration of the moving average is included, a VARMA(1,1) model best captures the DSC market and a VARMA(2,1) model best captures the VC market. We use the state-space representation of each of these models to obtain one-week-ahead aggregate market demand forecasts for each respective market.

Dynamic factor model

A Dynamic Factor Model is a multivariate model that seeks to identify one or several factors that are responsible for the co-movement of several time series data. For example, in the context of this study, a Dynamic Factor Model could be used to extract a single factor-the aggregate market demand for DSCs or VCs-, that is responsible for the co-movement of weekly sales and search volumes of DSCs or VCs. In the process of doing so, the Dynamic Factor Model can be used as a tool for obtaining short-term forecasts of each of these time-series data. (Stock et al., 2002; Gianonne et al., 2008).

To this end, we have explained how we used a variety of univariate and multivariate state-space models to obtain one-week-ahead forecasts of aggregate market demand in the DSC and VC markets. In what follows, we shall explain another technique that we used to make forecasts of sales volumes-the Long Short-Term Memory network, or LSTM for short.

**LSTM**

LSTMs are a kind of Recurrent Neural Network (RNN) that can learn long-term dependencies. First introduced by Hochreiter and Schmidhuber (1997), they have been refined and improved by many researchers over the years. An LSTM learns dynamically whether a given output should be retained and used as the next recursive input. Although many different versions of LSTMs have been developed, a basic LSTM typically involves a memory cell that can maintain its state over time.

The memory cell consists of an explicit memory (also known as the cell state vector) and gating units that regulate the flow of information into and out of the memory. Gates are sigmoid layers followed by pointwise multiplication operators. They are controlled by a concatenation of the output from the previous time step, the current input, and optionally the cell state vector. Typically, an LSTM has a forget gate that controls what information to discard from memory, an input gate that controls what new information is to be added to the cell state from current input, and an output gate that conditionally decides what to output from the cell state (Laddad, 2019).

Figure 9 shows how data flows through a memory cell and is controlled by each gate.

For each of the markets of our study, we construct 2 LSTM models-a univariate LSTM model that uses past values of sales volumes, and a multivariate LSTM model that uses past values of sales and search volumes to predict one-week-ahead sales volumes. We then compare the performance of the univariate model against that of the multivariate one, to determine if the use of search data (in addition to sales data) helps improve the accuracy of the model.

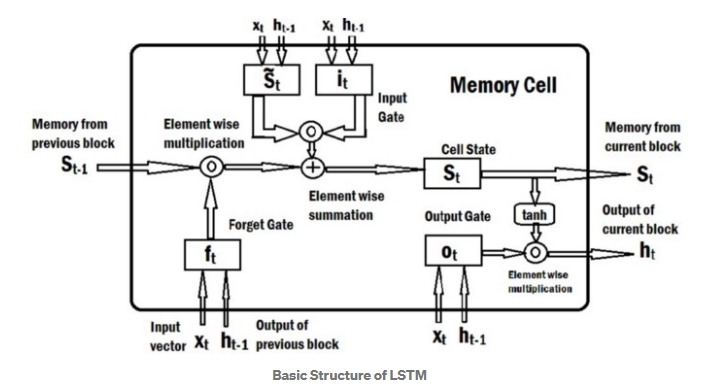


Figure 9 Graphical representation of the basic structure of an LSTM (Laddad, 2019).

**Results**

In the previous section, we explained the various state space and LSTM models we constructed to test the hypothesis that a multivariate model, which exploits the correlation between internet search volume and product sales volume, can outperform a univariate model based on historical product sales volume only.

With regards to the construction of multivariate models, it is worth revisiting a point that was raised in a previous chapter of this paper-the consideration of a possible time lag between sales and search data. As previously mentioned, it is logical to assume that sales data exhibits a lagged response vis-à-vis search data, since most customers would engage in information search for quite some time before purchasing the products of their interest. Based on the results of the Johansen cointegration test, we find that a 10-week lagged response of sales data with respects to search data for the DSC market, and a 7-week lagged response of sales data with respects to search data for the VC market leads to the highest rejection rate of the Johansen test’s null hypothesis.

To discern whether the consideration of the time lag between sales and search data is useful in improving the performance of a predictive model, we compare the performance of a multivariate LSTM model that does not consider the lagged response of sales data with respects to search data, against the performance of a multivariate LSTM model that does. In the latter model, search data is shifted with respects to sales data in the manner illustrated in Figure 5. We adopt the mean absolute percentage error (MAPE) as a metric of each model’s performance. Tables 7 and 8 compare the MAPE-s of each of these models, for each of the respective markets of our interest. The tables show that including the consideration of a time lag between sales and search data does indeed improve the performance of a multivariate LSTM model. For both the DSC market and the VC market, the model that considers the time lag has a lower MAPE than the model that does not.



Table 7 Comparison of the performance of a multivariate LSTM that does not consider the lagged response of sales data to search data versus a multivariate LSTM that considers a 10-week lagged response of sales data to search data, for the DSC market.



Table 8 Comparison of the performance of a multivariate LSTM that does not consider the lagged response of sales data to search data versus a multivariate LSTM that considers a 7-week lagged response of sales data to search data, for the VC market.

Based on the above finding, we include the consideration of a lagged response of sales data with regards to search data, for all the multivariate models discussed below. Shifting search data by 10 weeks for the DSC market and by 7 weeks for the VC market implies that the total number of weeks available for modelling differs between the 2 markets. We split sales and search data for each market into train and test sets, ensuring that each test set is of the same length, that is, that the test set for each market comprises the last 20 weeks in the data.

With regards to LSTM models, we conduct parameter tuning on the train set by creating models that exhaust the combinations of the values specified below, for the parameters:

-number of hidden layers: 1, 3, 7, 15

-batch size: 1, 5, 15, 30

-Regularization parameter: 0, 0.0001, 0.001, 0.01

-Number of lookback periods: 1, 2, 3, 5.

We then choose the model that has the lowest MAPE on the test set. The number of epochs is set to 1000 for each LSTM model.

For each of the models we construct, we adopt a direct multi-step forecast strategy, where we train a one-step model to predict the next value, based on all the **actual** observed values up to and including the current time step. This is opposed to a recursive multi-step forecast strategy, where a model is trained to predict the next value and the predicted value is appended to the end of the exogenous values fed into the forecast and used to predict future values (Brownlee, J. 2017). Each model’s performance is evaluated by the MAPE of the model on the test set.

Table 9 summarizes the performance of the models we constructed for the DSC market, while table 10 summarizes the performance of the models we constructed for the VC market. From these tables, we see that for both markets, the model with the best performance (or lowest MAPE) is the multivariate LSTM. This finding helps to answer our research question of whether a multivariate model, which exploits the correlation between internet search volume and product sales volume, can outperform a univariate model based on historical product sales volume only. Comparing the performance of the univariate LSTM with the multivariate LSTM, we see that combining search data with sales data can lower the MAPE of the univariate LSTM by about 28% (0.036 percentage points) for the DSC market and by about 8% (0.011 percentage points) for the VC market.[[4]](#footnote-4) A graphical representation of the multivariate LSTM for each respective market is provided by Figures 10 and 11. In each of these figures, the black line represents the trajectory of observed weekly sales volumes (in JPY), the green line represents predicted weekly sales volumes on the train set and the red line represents predicted weekly sales volumes on the test set.

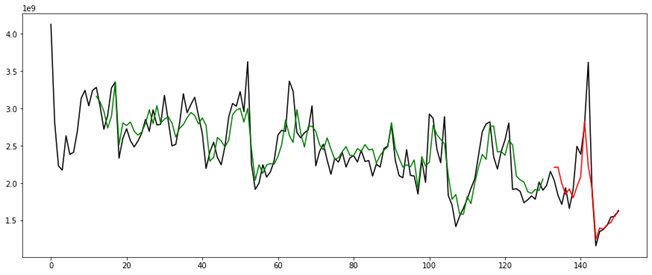


Figure 10 Multivariate LSTM model for the **DSC market**, where the black line represents the trajectory of observed weekly sales volumes (in JPY), the green line represents predicted weekly sales volumes on the train set and the red line represents predicted weekly sales volumes on the test set.

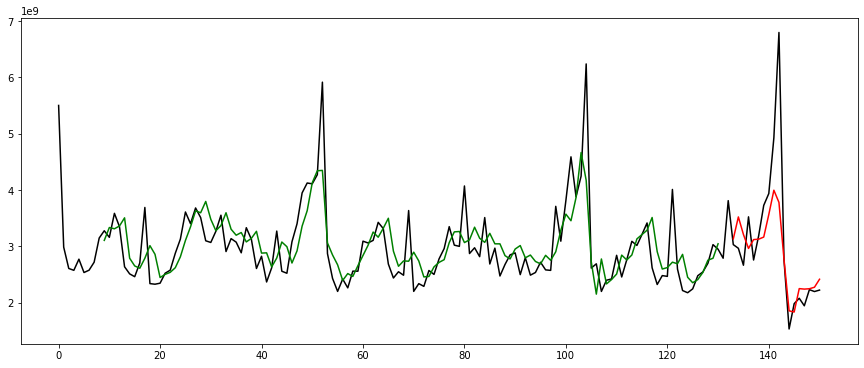


Figure 11 Multivariate LSTM model for the **VC market**, where the black line represents the trajectory of observed weekly sales volumes (in JPY), the green line represents predicted weekly sales volumes on the train set and the red line represents predicted weekly sales volumes on the test set.

Comparing the performance of multivariate state-space models against the performance of univariate state-space models, we see that with regards to the DSC market, multivariate autoregressive models (i.e. VAR and VARMA) outperform univariate unobserved components models (i.e. the local level, local linear trend, local level model with seasonality and local linear trend model with seasonality), but not necessarily univariate autoregressive models. For example, the MAPE of the VARMA model is 0.190, which is higher than the MAPE of the ARIMA and SARIMA models-0.170 and 0.176, respectively. In addition, we see that the multivariate Dynamic Factor Model outperforms all univariate state space and autoregressive models.

With regards to the VC market, we see from Table 10 that multivariate autoregressive models (i.e. the VAR and VARMA models) and the Dynamic Factor Model outperform all the univariate models, except for the univariate LSTM.

The finding that multivariate state-space models can perform better in terms of prediction than some univariate state space models provides further evidence in support of our claim that combining search data with sales data helps improve the performance of a predictive model.



Table 9 Performance of univariate and multivariate models on the test set of the data for the DSC market.



Table 10 Performance of univariate and multivariate models on the test set of the data for the VC market.

**Discussion**As we explained in the introductory section of this paper, previous works on retail demand forecasting have not considered whether a multivariate model, which exploits the correlation between internet search and product sales volumes, can outperform a univariate model based on product sales volumes only. To fill this gap in the literature, we construct several state space and LSTM models for the DSC and VC markets in Japan and compare the performance of univariate models against that of multivariate models. Our findings contribute to the existing literature in 3 specific ways.

First, they serve as evidence that using internet search data in addition to historical sales data can help to improve the performance of a model designed to make short-term forecasts of market demand/sales. For example, by comparing the performance of univariate state space and LSTM models against that of multivariate state space and LSTM models, we find that the multivariate LSTM performs the best for both the DSC and VC market. In addition, we discover that for the DSC market, multivariate state-space models, namely VAR and VARMA, can outperform some univariate state-space models, namely the unobserved components class of models. Likewise, we discover that for the VC market, multivariate state-space models, namely the VAR, VARMA and Dynamic Factor Model, can outperform all the univariate models except for the univariate LSTM.

The second contribution of our paper is its attempt to not only highlight but also quantify, the effectiveness of the use of internet search data as an explanatory variable in the short-term forecasting of future demand or sales volumes. To be more specific, we discover that combining search data with sales data can lower the MAPE of an LSTM model by about 28% (0.036 percentage points) for the DSC market and by about 8% (0.011 percentage points) for the VC market.

Finally, the third contribution of this paper is its attempt to answer the research question: can highly accurate short-term forecasts of consumer demand be achieved via relatively novel techniques, such as state-based modelling and LSTM? We provide a positive answer to this question by achieving high model performance, or a low MAPE on the test set of our data. For example, with regards to the DSC market, we show that a multivariate LSTM model can achieve a MAPE as low as 0.092 and that a Dynamic Factor Model can achieve a MAPE as low as 0.167 on the test set. Similarly, with regards to the VC market, we show that a multivariate LSTM model can achieve a MAPE as low as 0.126 and that a VAR Model represented in state- space can achieve a MAPE as low as 0.194 on the test set.

Having said so, a limitation of our approach is that it can only produce highly accurate forecasts over a short-term horizon such as one, or at best, a few weeks ahead. As mentioned in an earlier section of this paper, we adopt a direct multi-step forecast strategy, where we train a one-step model to predict the next value, based on all the **actual** observed values up to and including the current time step, as opposed to a recursive multi-step forecast strategy, where a model is trained to predict the next value and the predicted value is appended to the end of the exogenous values fed into the forecast and used to predict future values (Brownlee, J. 2017).

The reason for our choice of approach is that we want to avoid a situation where prediction errors accumulate over time, causing the performance of the model to quickly degrade as the prediction time horizon increases. Such an approach allows us to achieve highly accurate short-term forecasts at the expense of not being able to make forecasts over a long-term horizon.

**Conclusions**

In this research, we strived to discover if more accurate forecasting of sales or product demand could be achieved by combining internet search data with POS (/sales volume) data. The research questions that we sought to answer are as follows.

(1) Is search term volume a good measure of consumer demand for a retail good?

(2) How can the use of this measure, in combination with data on sales volumes, reduce the error (or improve the accuracy) of a predictive model?

(3) Can highly accurate short-term forecasts of consumer demand/sales volume be achieved via relatively novel techniques, such as state-based modelling and LSTM?

To answer these questions, we first perform the Granger causality test and Johansen cointegration test on search and sales volume data, to better understand the statistical relationship between them. The results of these tests indicate that search and sales data are useful in forecasting each other and that these time series data share a cointegrating relationship.

Next, we consider a possible lagged response of sales volume data with respects to search volume data. This is because most customers might engage in information search for quite some time before purchasing the products of their interest. To discern whether the consideration of the time lag between sales and search data is useful in improving the performance of a predictive model, we compare the performance of a multivariate LSTM model that does not consider the lagged response of sales data with respects to search data, against the performance of a multivariate LSTM model that does. The result of this experiment shows that for both the DSC market and the VC market, the model that considers the time lag has a lower MAPE than the model that does not. More precisely, we find that a 10-week lagged response of sales data with respects to search data for the DSC market, and a 7-week lagged response of sales data with respects to search data for the VC market leads to the highest rejection rate of the Johansen test’s null hypothesis. Based on these findings, we include the consideration of a time lag when constructing multivariate state space and LSTM models for making short-term predictions of aggregate demand or sales.

To address the question of whether combining search data with sales data can help to improve the performance of a predictive model, we construct 4 univariate state-space models, where historical sales volumes are used to make one-week-ahead forecasts of aggregate market demand, and 2 multivariate state-space models, where historical sales and search volumes are used to make one-week-ahead forecasts of aggregate market demand, for each of the markets of our interest. By comparing the performance of univariate models against that of multivariate models, we find that the multivariate LSTM performs the best for both the DSC and VC market. We also find evidence to support the claim that multivariate state-space models have the potential to perform better than some univariate ones.

As far as we are aware, this is the first paper in the literature that addresses all 3 of the research questions we raised. It is also the first that makes use of state-space modelling and LSTMs to forecast future demand for DSCs and VCs in the Japanese market. However, a limitation of our approach is that it can only produce highly accurate forecasts over a very short-term horizon such as one, or at best, a few weeks ahead. As an extension to this paper, it is worth exploring how search data can be combined with sales data to produce relatively accurate forecasts of demand/sales across longer time horizons, such as a quarter of a year.

**Acknowledgements**

We would like to express our gratitude to GFK Japan for the provision of the search and sales data that we used in this study, and for permitting us to publish the results of our research. GFK Japan had no role in the preparation of this manuscript, or the decision to publish it.

**References**

Albert Analytics Technology. (n.d.) 「状態空間モデルとは」.https://www.albert2005.co.jp/knowledge/machine\_learning/ssm/about\_ssm

Alon I., Min Q., Sadowski R.J. (2001). Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional method. *J. Retail. Consum*, *Serv* 8(3), 147-156.

Askitas, Nikos and Zimmerman, Klaus F. (2009). Google econometrics and unemployment forecasting. IZA Discussion Papers 4201, Institute of Labor Economics (IZA).

Bandara K., Shi P., Bergmeir C., Hewamalage H., Tran Q., Seaman B. (2019). Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology. In Gedeon T., Wong K., Lee M. (eds) *Neural Information Processing. ICONIP 2019. Lecture Notes in Computer Science, vol 11955*, Springer, Cham. <https://doi.org/10.1007/978-3-030-36718-3_39>

Borovykh, A., Bohte, S. & Oosterlee, C.W. (2017). Conditional time series forecasting with convolutional neural networks. arXiv [cs.AI].

Brownlee, Jason. (2019). *4 strategies for multi-step time series forecasting*. Machine Learning Mastery. <https://machinelearningmastery.com/multi-step-time-series-forecasting/>

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665-676.

Hochreiter, S., Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In *Advances in neural information processing systems 9*. Cambridge, MA: MIT Press.

Kholodilin A. Konstantin, Podstawski Maximilian, and Siliverstovs Boriss (2010). Do Google searches help in nowcasting private consumption? A real-time evidence for the US. DIW Berlin Discussion Paper No. 997, German Institute for Economic Research.

Koller D, Friedman N. (2009). *Probabilistic Graphical Models*. Cambridge, MA: MIT Press.

Kotzé Kevin. (n.d.) *State-Space Modeling*. https://kevinkotze.github.io/ts-4-state-space/.

Laddad, Amey. (2019). Basic understanding of LSTM. *Good Audience*. https://blog.goodaudience.com/basic-understanding-of-lstm-539f3b013f1e

Meyer, Robert J. (1982). A descriptive model of consumer information search

Behavior. *Marketing Science,* 1(1), 93–121.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997). Consumer

information search revisited: Theory and empirical analysis. *Journal of*

*Consumer Research,* vol. 23 (March), 263–277.

Preiss Tobias, Reith Daniel, Stanley H. Eugene (2010). Complex dynamics of our economic life on different scales: Insights from search engine query data. *Philosophical Transactions of the Royal Society A*, 368, 5707-5719.

Ramos Patricia, Santos Nicolau, and Rebelo Rui (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-Integrated Manufacturing* 34, 151-163

Salinas David, Flunkert Valentin, Gasthaus Jan, and Januschowski Tim (2019). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting* 36(3), 1181-1191.

Quantstart. (n.d.) *State Space Models and the Kalman Filter*. https://www.quantstart.com/articles/State-Space-Models-and-the-Kalman-Filter/

Rothenberg, T. (2007). State Space Models and the Kalman Filter. https://eml.berkeley.edu/~rothenbe/Fall2007/kalman.pdf

Stock, J.H. and Watson, M.W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economics Statistics*, 20, 147-162.

Urbany, Joel E., Peter R. Dickson, and William L. Wilkie (1989). Buyer uncertainty

and information search. *Journal of Consumer Research,* 16(2), 208–215.

Wu Lynn and Erik Brynjolfsson (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker (Ed.s), *Economic Analysis of the Digital Economy* (Chapter 3, pp.89-118). Illinois: University of Chicago Press.

Zhang, Mingda (2018). Time Series: Autoregressive models AR, MA, ARMA, ARIMA. https://people.cs.pitt.edu/~milos/courses/cs3750/lectures/class16.pdf

1. GFK. (2021, 5, 4). In Wikipedia. https://en.wikipedia.org/wiki/GfK [↑](#footnote-ref-1)
2. Traditional sales refer to all sales transactions made in stores while internet sales refer to all sales transactions made via e-commerce websites. [↑](#footnote-ref-2)
3. The aggregation of the internet search data was conducted by Yahoo Japan Corporation. [↑](#footnote-ref-3)
4. The formula for percentage change is:

   . [↑](#footnote-ref-4)