WORKING PAPER

ITLS-WP-21-21

Forecasting Automobile Gasoline Demand in Australia Using Machine Learning-based Regression

By
Zheng Li[a], Bo Zhou[a] and
David A. Hensher[b]

[a] School of Economics and Finance,
Xi'an Jiaotong University, Shaanxi 710049, PR China

[b] Institute of Transport and Logistics Studies (ITLS),
The University of Sydney, Australia

November 2021

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-21-21

TITLE:

Forecasting Automobile Gasoline Demand in Australia Using Machine Learning-based Regression

ABSTRACT:

We use a variant of machine learning (ML) to forecast Australia's automobile gasoline demand within an autoregressive and structural model. By comparing the outputs of the various model specifications, we find that training set selection plays an important role in forecasting accuracy. More specifically, however, the performance of training sets starting within identified systematic patterns is relatively worse, and the impact on forecast errors is substantial. Instead of treating these patterns as noise, we explain these systematic variations in machine learning performance, and explore the intuition behind the 'black-box' with the support of economic theory. An important finding is that these time points coincide with structural changes in Australia's economy. By examining the out-of-sample forecasts, the model's external validity can be demonstrated under normal situations; however, its forecasting performance is somewhat unsatisfactory under event-driven uncertainty, which calls on future research to develop alternative models to depict the characteristics of rare and extreme events in an ex-ante manner.

KEY WORDS: ***Energy demand forecasting; machine learning; time series; structural changes; automobile sector***

AUTHORS: **Li, Zhou and Hensher**

Acknowledgements: This work was funded by the Young Talent Support Plan, Xi'an Jiaotong University [0002019060, 2019]. We thank three referees and the subject editor for comments.

CONTACT:

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H04)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1813

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: November 2021

1. Introduction

Petroleum fuels have played a vital role in the contemporary business world, supporting many economic activities such as manufacturing, agriculture and trade. To date, such fuels remain the primary energy source for realising mobility and accessibility through transport systems, in particular the automobile sector, although we recognise the growing commitment to electric vehicles. While it might be expected that prices of all fuel sources would adjust to reflect supply and demand, there is an expectation that fossil fuels will remain the predominant source of energy for passenger cars over the next few decades as societies prepare the ground for substitution into clean or cleaner fuels. Energy demand forecasting is pivotal to planning, decision making and formulating energy policies, which relies on econometric/statistical methods (e.g., Ediger and Akar 2007; Li *et al.* 2010) and, more recently, machine learning (ML) methods (e.g., Forouzanfar*et al.* 2012; Beyca *et al.* 2019). The former explicitly establish the relationship between energy consumption and its influencing factors, which requires certain pre-specified assumptions (e.g., the functional form and statistical distribution of parameters).

Machine learning replicates human learning through algorithms and forecasts the outcome variables (y) given some other variables, x, which are *features* in the language of ML or *independent variables* in the language of econometrics. A major limitation of the current literature on the application machine learning to energy is that some papers emphasize the computer science perspective optimizing computational parameters such as the accuracy rate while the economic or finance intuition might be ignored (Mullainathan and Spiess 2017; Ghoddusi *et al.* 2019). For example, data that diminishes the accuracy rate would be treated as noise, and the potential economic implications are often ignored. Meanwhile, other papers concentrate on the economic or financial perspective without fully exploring the capacity of the algorithms to solve or explain the problem under study (Ghoddusi *et al*. 2019, p. 720), which has low calculation efficiency. A recent assessment by Allen (2019) promotes the view that machine learning may accelerate a crisis in science given that machine learning algorithms have been developed specifically to find interesting things in datasets, and so when they search through huge amounts of data, they will inevitably find a pattern.

In this paper, we systematically investigate the role of ML techniques, model features and training periods in forecasting Australia's automobile gasoline demand, and present evidence on the forecasting power of a model framework driven by machine learning in the case of a small sample. In this situation, we paygreater attention to the selection of training sets when using machine learning methods in order to minimise the estimation error caused by the unique characteristics of the small sample. Our empirical findings below suggest that when the forecasts are systematically influenced by one subset of training data (covering the period of 1992-1997), the impact on forecast errors is significant. We have found that these time points coincide enabling us to interpret the identified systematic variation in ML performance and its economic intuition, which is typically absent in the literature. We also assess the out-of-sample forecasts of the selected optimal model, its external validity can be demonstrated under stable conditions; However, its forecasting performance is somewhat unsatisfactory under event-driven uncertainty (i.e., Corona Virus Disease 2019 (COVID-19) in this study), which calls on future research to develop new models to anticipate the characteristics of extreme events.

2. Forecasting Transport Energy Demand Using ML Techniques

Machine learning, proposed by Samuel (1959), has been widely used in the fields of energy and economics with diverse applications, for example, the optimization of energy inputs (Nabavi-Pelesaraei *et al*., 2014; Abdelaziz *et al*., 2016; Nabavi-Pelesaraei *et al*., 2017; Ali & Abd Elazim, 2018; Khanali *et al*., 2021), the investigation of energy efficiency (Nabavi-Pelesaraei *et al*., 2014), forecasting energy commodity prices (Ding, 2018; Yu *et al*., 2017; Zhang et al., 2015), forecasting energy demand (Yang *et al*., 2014; Panapakidis and Dagoumas, 2017; Ou *et al*., 2020; Haque *et al*., 2021). Popular ML techniques in the relevant literature include applied artificial neural networks (ANN) (Olanrewaju et al., 2013; Kunwar et al., 2013); deep learning (Lago et al., 2018; Peng et al., 2018), support vector machine (SVM) (Papadimitriou et al., 2014; Zhu *et al*., 2016; Jiang *et al*., 2018), decision trees (Bastardie *et al*., 2013; Zhao and Nie, 2020) and ensemble methods (Ghasemi et al., 2016; Mirakyan et al., 2017).

Given our research focus, only studies that have used ML to forecast transport energy demand are reviewed, as summarised in Table 1 (For a broader review on ML applications in forecasting various energy types such as crude oil, natural gas and power and for different purposes such as price forecasting and data management, please refer to Ghoddusi *et al.* (2019). Earlier-published papers adopted one type of ML technique per study to forecast transport energy demand. Haldenbilen and Ceylan (2005) used a genetic algorithm (GA) for Turkey's annual transport energy demand forecasting; while Murat and Ceylan (2006) and Kazemi *et al.* (2010) and Limanond *et al*. (2011) applied artificial neural networks for Turkey, Iran and Thailand respectively. More recent work has developed multiple energy demand forecasting models with various ML techniques, and there is mixed evidence on the optimal machine learning method, which may be data specific. For example, Forouzanfar *et al.* (2012) proposed the multi-level genetic programming (MLGP) approach to forecast transport energy demand in Iran, and found that its forecasting accuracy is similar to that of neural network. Teng *et al.* (2017) forecasted China's total transport energy demand, using ANN, the group method of data handling (GMDH) and the support vector machine. Their empirical results show that GMDH delivered more accurate forecasts than other ML techniques. Azadeh *et al.* (2015) predicted Iran's weekly gasoline demand, and found that the support vector regression (SVR) approach outperformed ANN.

Several studies have assessed the role of feature selection; for example, with five candidate features (Gross domestic product (GDP), population, oil price, the number of vehicles, and passenger transport volumes), Geem (2011) considered four different feature combinations when forecasting transport energy consumption in South Korea. They found that the approach that accounted for the number of vehicles and the oil price resulted in the optimal solution and best forecasting performance. Teng *et al.* (2017) found that the urbanisation rate is an important feature for forecasting China's transport energy demand. In each reviewed study, a fixed dataset was used for training their machine learning models, without training data selection. An examination of Table 1 also suggests that the most common features used in the ML models for forecasting transport energy are national income, population, the number of vehicles, and fuel price.

Table 1: Transport energy demand forecasting studies using ML techniques

Reference	Location	Data period	ML techniques	Model features as predictors	ML technique selection	Feature selection	Training set section
Haldenbilen and Ceylan (2005)	Turkey	1970-2000 (annual)	Genetic algorithm	GDP, vehicle kilometers traveled, the number of vehicles	No	Yes	No
Murat and Ceylan (2006)	Turkey	1970-2001 (annual)	ANN	Gross national product (GNP), population and vehicle kilometers	No	No	No
Kazemi *et al.* (2010)	Iran	1968-2007 (annual)	ANN	GDP, population, and the number of vehicles	No	Yes	No
Limanond *et al.* (2011)	Thailand	1989-2008 (annual)	ANN	GDP, population, and the number of vehicles	No	Yes	No
Teng *et al.* (2017)	China	1980-2011 (annual)	ANN, SVM, GMDH	Disposable income, population, vehicle registrations, fuel price index, and urbanisation rate	Yes	Yes	No
Azadeh *et al.* (2015)	Iran	Aug. 2009 to Dec. 2011 (weekly)	ANN, SVR	Transported freight per kilometre, transported passengers per kilometre, and the number of holidays per week	Yes	No	No
Geem (2011)	South Korea	1990-2007 (annual)	ANN, Multiple linear regression(MLR)	GDP, population, oil price, the number of vehicle registrations, passenger transport amount	Yes	Yes	No
Forouzanfar *et al.* (2012)	Iran	1968-2005 (annual)	Genetic programming, ANN	GDP, population, and the number of vehicles	Yes	No	No

3. Forecasting Methods

For a training data set $D = \{\boldsymbol{x}, y\}$, we define $\boldsymbol{x}$ as a vector of independent variables or features, and y as the outcome variable. A linear regression function designed to identify a hyperplane is given in equation (1), estimated by minimising a loss function: $\min||y - f(x)||^2$, usually referred to as the empirical risk minimisation function (Vapnik 1998).

$$y = f(\boldsymbol{x}) = \boldsymbol{w}^T x + b \tag{1}$$

By adding a l_1-norm regularisation term, we obtain the least absolute shrinkage and selection operator (LASSO) regression model proposed by (Tibshirani 1996):

$$\min||y - f(\boldsymbol{x})||^2 + \lambda||\boldsymbol{w}||_1 \tag{2}$$

where $\lambda \geq 0$ is a tuning parameter, and for λ=0 it reduces to an ordinary least square (OLS) regression. For details on LASSO and its applications in the field of energy demand, see Besagni and Borgarello (2018), Mashhadi and Behdad (2018) and Guo *et al.* (2018).

The support vector regression (SVR) algorithm, developed by Vapnik *et al.* (1992) and Cortes and Vapnik (1995), is an extension of the support vector machine and has gained popularity in the field of energy economics (see Ghoddusi *et al*. 2019 for a survey). The major advantage of SVR over other ML techniques is that it results in a convex minimisation problem with a unique global minimum, which avoids local minima (Plakandaras *et al.* 2017). SVR adopts the structural risk minimisation (SRM) principle by seeking to minimise an upper bound of the generalisation error rather than the training error, which results in improved generalisation performance, the absence of a local minimum and the sparse representation of a solution. The SVR model defines a margin ϵ for the hyperplane, and the loss function of a SVR model can be written as:

$$\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} l_\epsilon \left(f(x_i) - y_i\right) \quad \text{s.t. } l_\epsilon(z) = \begin{cases} 0, & \text{if } |z| \le \epsilon \\ |z| - \epsilon, & \text{otherwise} \end{cases} \tag{3}$$

ML techniques have been primarily used to pursue the quality of predictions (Yu *et al.* 2012; Georges and Pereira 2020). With respect to the application of machine learning, a critical challenge is that it should be guided by some sensible quantum of economic and behavioural theory (Mullainathan and Spiess 2017). This current paper attempts to shed some light on the identified systematic variations in ML performance from an economics perspective, using Australia's automobile gasoline demand as the empirical application, introduced in the following section.

4. Data

In Australia, around 85.5% of passenger cars use gasoline as the source of combustion; while 12.8% use diesel and 1.7% consume other types of fuel such as Liquefied Petroleum Gas (LPG) and electricity (ABS 2020). This study concentrates on gasoline demand, given its dominant role in Australia's automobile sector. To forecast Australia's total automobile gasoline consumption (TAGC) in million litres, we compiled a quarterly time series database spanning the period from Quarter 1 of 1974 (1974Q1) to Quarter 2 of 2019, from various sources, mainly the Australian Bureau of Statistics, Department of Industry, Tourism and Resources (Australia) and the Department of the Environment and Energy (Australia). The variables serving as predictors include: real gasoline price (*RGP*) in Australian dollars (Au$), Australia's real household gross disposable income (*RHGDI*), population in millions (*POP*), expenditures on hotels, cafes and restaurants (*HCR*) in million Au$, final consumption expenditure (*FCE*) in million Au$, and purchase of vehicles (*POV*) in million Au$. Table 2 summarises the descriptive statistics of the time series with Australian automobile gasoline consumption shown in Figure 1 from 1974Q1 to 2019Q2 .

Table 2: Descriptive statistics

	TAGC	*POP*	*RHGDI*	*RGP*	*FCE*	*POV*	*HCR*
Mean	4201	18.37	228508	1.34	138175.74	2875	10635

Median	4382	18.04	193284	1.33	117867.0	2107	9265
Standard deviation	581	3.48	94624	0.22	62818.90	1593	3613
Minimum	2596	13.01	93942	0.85	54004.0	1033	5665
Maximum	5243	25.41	410253	1.92	272594.0	6288	18647

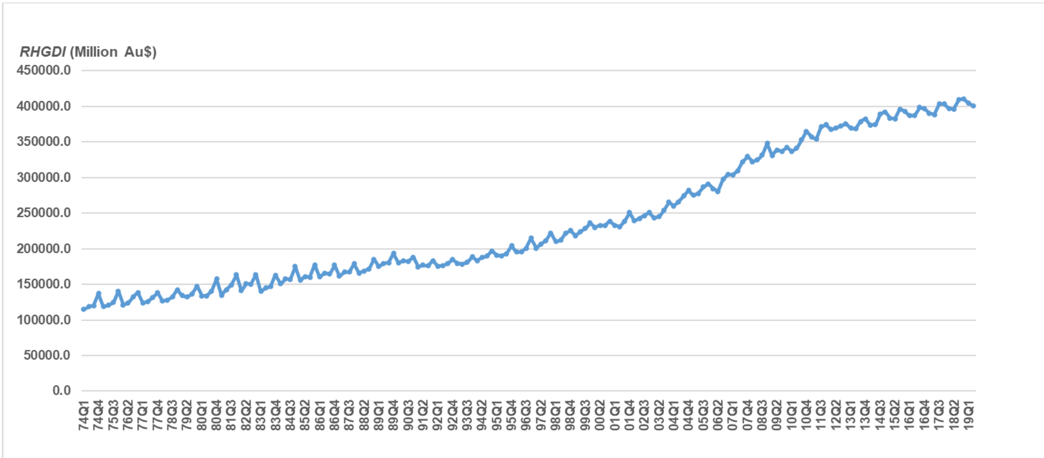


Figure 1: Australian automobile gasoline consumption from 1974Q1 to 2019Q2

The full sample size of our data is 182. We divided the data into two parts. The first section (1974Q1-2017Q2) is used for training models. To investigate the role of training set selection, we varied the starting point of the training set from 1974Q1 to 2015Q1. The full training dataset has the maximum number of observations of 174 (i.e., 1974Q1-2017Q2); while the smallest size is 10 (i.e., 2015Q1-2017Q2). To avoid overfitting (see Barutcuoglu *et al*. 2006), a three-fold cross-validation is applied to determine the optimal parameter combination of the ML model. The training data is divided into three subsets. For each parameter combination, two subsets of the data are used to train the model and the rest for validating the model performance. This process is repeated three times for each combination, and the average performance measures across all parameter settings (see Table 3 below) are calculated. The parameter combination with the best average performance measure is selected as the optimal model. The remaining observations, from 2017Q3 to 2019Q2, are used as a hold-out sample to examine the forecasting performance of ML models with different training periods and features. This flexible treatment allows us to identify the key influences on forecasting accuracy: the type of ML model, the length of training period or the combination of model features.

It is typically assumed that transport gasoline consumption increases with income (Dahl and Sterner 1991). We plotted the relationship between income and gasoline demand in Figure 2, which shows clear nonlinearity. For *RHGDI* less than Au$ 200,000 million, gasoline consumption increased with income. When *RHGDI* is between $Au200,000m and $Au300,000m, this increasing trend weakened, and then gasoline consumption tended to slightly decrease with income when *RHGDI* surpassed $Au300,000m (i.e., after the real gasoline price peaked in the year of 2008). Can these figures imply that Australians have travelled less over the recent period? In 2018, Australia's passenger cars travelled 179,761 million kilometres (ABS 2019), more than the total distance of 157,935 million kilometres in 2007 (ABS 2008). The improvement of fuel economy may play a role in this transition (i.e., longer while more fuel-efficient travel), according to the Bureau of Infrastructure, Transport and Regional Economics (BIRE 2017).

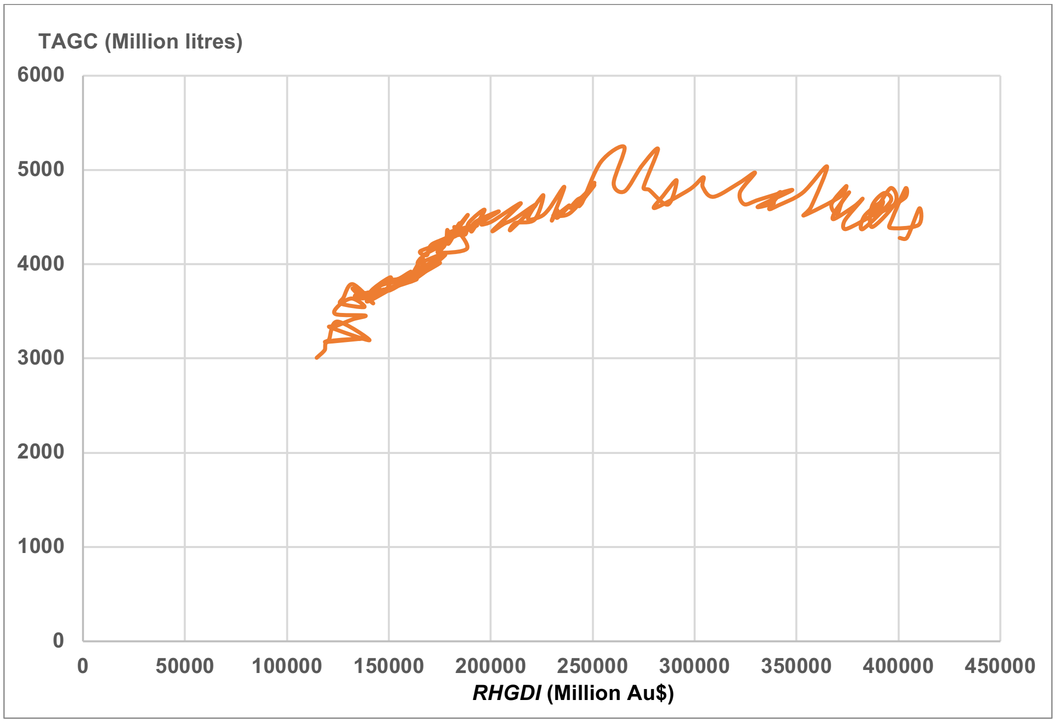


Figure 2: Nonlinear relation between gasoline consumption and income

5. Model Specification

Having introduced the time-series data, the general model specification for the empirical simulations is given in equation (4).

$$y_{t+h} = \sum_{k=1}^{m} a_{1,k} y_{t\text{-}k} + \sum_{k=1}^{m} a_{2,k} Basic_{t\text{-}k} + \sum_{k=1}^{m} a_{3,k} POV_{t\text{-}k} + \sum_{k=1}^{m} a_{4,k} HCR_{t\text{-}k} + \sum_{k=1}^{m} a_{5,k} FCE_{t\text{-}k} \tag{4}$$

y is the outcome variable *TAGC*; m represents the number of lags for the system; h is the forecasting horizon in quarters with three alternative horizons (h) considered:1, 2, and 4 quarters ahead. National income, population, the number of vehicles, and fuel price were frequently used as predictors in existing ML models for transport energy demand. Given that the number of Australian motor vehicles is not fully available during all of the sample period, this feature has not been included in the forecasting models. Therefore, national income, population and gasoline price are the *Basic* features for the ML models. In total, the model features consist of three parts: the autoregressive (AR) component (y_{t-k}), the *Basic* features consisting of *RHGDI*, *POP* & *RGP*, and three additional features: *HCR, POV, & FCE*. For time-series demand forecasting, there is no need to pre-process the raw data, given that seasonality, unit-root or other characteristics are treated as additional features, which are incorporated into the final ML algorithm for forecasting (Ghoddusi *et al*. 2019).

The forecasting performance of alternative model forms is evaluated using the mean absolute percentage error (MAPE) criterion (see equation 15). There are various error measures such as the mean square error (MSE) and the mean absolute error (MAE). By expressing it as a percentage error (normalised by actual demand), the MAPE allows for a meaningful comparison across datasets and/or across studies; however, other absolute indicators without such a normalisation can be associated with different magnitudes. The value of MAPE provides intuitive information on how accurate the forecast is; for example, MAPE≤0.1 indicates a percentage error of no more than 10%, suggesting a high level of accuracy (Lewis 1986).

$$\mathrm{MAPE} = \frac{\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right|}{n} \tag{5}$$

y_t is the observed gasoline consumption in time period t and $\widehat{y_t}$ is the corresponding forecast; n is the total number of the hold-out-sample observations (n=8 in this current study). All experiments are coded using the Python language and implemented by a server. The parameter domains of the LASSO and SVR models are given in Table 3.

Table 3: Parameter domains

Parameter	Description	Domain
λ	Weight of LASSO	{0.1, 0.3, 0.5, 0.7, 0.9}*{1e-2, 1e-1, 1e0, 1e1, 1e2}
C	Penalty of SVR	{0.1, 0.2, 0.5}*{1e-2, 1e-1, 1e0, 1e1, 1e2}
λ	Width of RBF kernel	{0.1, 0.2, 0.5}*{1e-2, 1e-1, 1e0, 1e1, 1e2}

6. Results and Discussion

6.1: Experiment 1 - Choosing the preferred ML Model using the full training data

Existing energy demand forecasting studies used their full training data to train their ML models, and the one with the best forecasting accuracy is chosen as the preferred forecasting model (see e.g., Forouzanfar *et al*. 2012; Tang *et al.* 2017). In this experiment, we followed this standard practice to determine the overall best model for the empirical application. Table 3 summarises the MAPE values across different models and feature combinations, trained by using the full training set: 1974Q1-2017Q2.

Table 3: Forecasting performance under different methods and feature combinations (Total training data: 1974Q1-2017Q2)

Model	Feature	Forecasting horizon (quarters)		
		1	2	4
OLS	AR+Basic	0.0258	0.0349	0.0441
	AR+Basic+HCR	0.0275	0.0377	0.0415

	AR+Basic+POV	0.0231	0.0338	0.0462
	AR+Basic+FCE	0.0258	0.0366	0.0439
	AR+All	0.0251	0.0339	0.0397
LASSO	AR+Basic	0.0241	0.0298	0.0371
	AR+Basic+HCR	0.0289	0.0334	0.0356
	AR+Basic+POV	0.0214	0.0292	0.0348
	AR+Basic+FCE	0.0270	0.0348	0.0371
	AR+All	0.0201	0.0293	0.0347
SVR-Linear	AR+Basic	0.0267	0.0364	0.0453
	AR+Basic+HCR	0.0238	0.0375	0.0517
	AR+Basic+POV	0.0241	0.0351	0.0529
	AR+Basic+FCE	0.0256	0.0357	0.0467
	AR+ALL	0.0205	0.0338	0.0521
SVR-RBF	AR+Basic	0.0201	0.0231	0.0268
	AR+Basic+HCR	0.0194	0.0222	0.0262
	AR+Basic+POV	0.0183	0.0215	0.0272
	AR+Basic+FCE	0.0187	0.0218	0.0254
	AR+ALL	0.0173	0.0215	0.0263

In Table 4, we report the hold-out-sample MAPE results for one-, two- and four-step-ahead forecasts respectively, using the full training set from 1974Q1 to 2017Q2. The overall performance is acceptable, with the lowest/highest percentage error being 1.73%/5.29%. Three linear methods (linear regression with OLS as the benchmark, LASSO and SVR with linear kernel) and one nonlinear method (SVR with radial basis function (RBF) kernel) are compared in terms of forecasting accuracy, and the results show that (1) except for the linear SVR, ML methods outperform the statsitical method; (2) overall, the nonlinear SVR-RBF delivers the best forecasting performance, with the greater improvement in accuracy for longer horizons. Table 4 also shows that, relative to the Basic specification with the autoregressive component only, *RHGDI*, *POP* and *RGP*, each additional feature has improved the performance of SVR-RBF, in which the role of *FCE* is the greatest among three additional features. Adding all of them simultaneously would lead to the best forecast for Horizon=1&2; while the inclusion of *FCE* is optimal for Horizon=4. Considering its overall performance and capability to address nonlinearity in the data, the SVR-RBF model is selected as the empirical model in this paper.

6.2 Experiment 2 - Identifying the role of training set selection and explaining its economic implications

This experiment investigates the stability of the preferred model (SVR-RBF) by varying the size of the training set from 10 (i.e., 2015Q1-2017Q2) to 174 (i.e., 1974Q1-2017Q2). The hold-out-sample forecasting performances of SVR-RBF with different feature combinations are shown in Figures 3-5 for Horizon=1, 2 & 4 respectively. Though having some fluctuations, their MAPE values (all below 0.0045) would suggest high levels of forecasting accuracy.

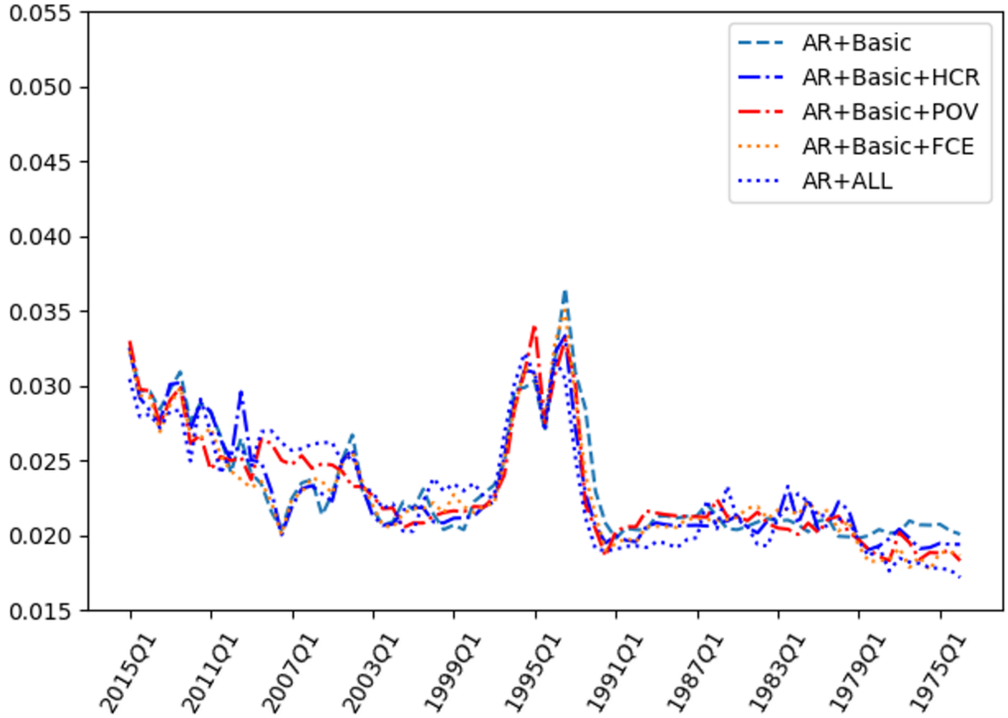


Figure 3: SVR-RBF MAPE values (y-axis) over different starting quarters (x-axis) and model features for Horizon=1

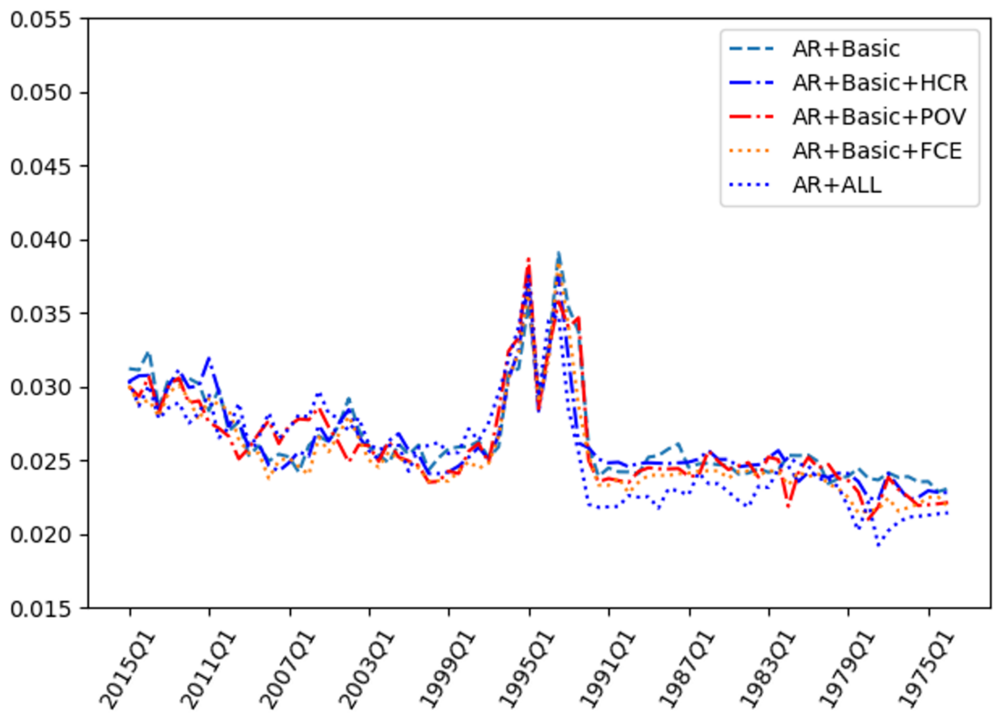


Figure 4: SVR-RBF MAPE values (y-axis) over different starting quarters (x-axis) and model features for Horizon=2

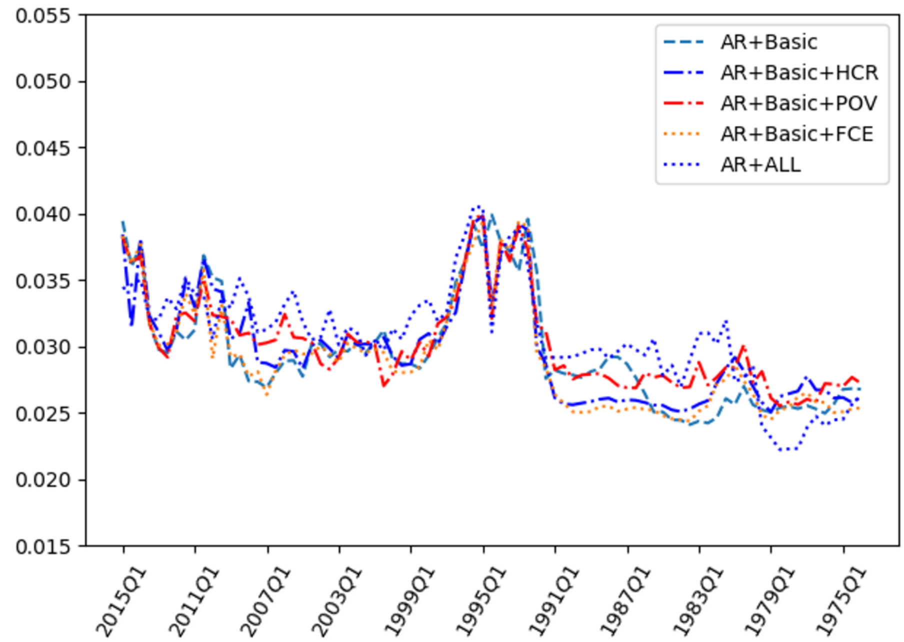


Figure 5: SVR-RBF MAPE values (y-axis) over different starting quarters (x-axis) and model features for Horizon=4

Given that the SVR-RBF model with all features delivered a slightly better overall forecast than other feature combinations, it is used as the descriptive example for demonstrating the role of training data selection. Training sets starting with 1974Q3, 1978Q1 & 1979Q1 are associated with the lowest percentage errors, namely 1.72%, 1.93%, and 2.22% for Horizon=1, 2 & 4 respectively, better than forecast by using the full training set (see Table 4). For the starting point between 1992 and 1997, there is a clear 'M' shape with double tops, with its key points summarised in Table 5. For those training sets which start within the left-right boundaries of 'M', their MAPE values are significantly greater than those of other sets, where the starting points being 1995Q3 and 1995Q1 have produced the least accurate forecasts for Horizon=1 and Horizon=2&4 respectively. Relative to the direct interpretation using the values in Table 5, a more informative approach is to use their relative differences, which demonstrate stronger inconsistency in forecasts. For example, holding all other factors constant, by shifting the starting point of the training set from the left peak to the left boundary of 'M', this reduction in training size would significantly reduce errors (e.g., from 3.22% to 2.26% for Horizon=1). Given the average error across all training sets of 2.31%, 2.57% and 3.09% for one-, two- and four-step-ahead forecasts respectively, Figures 3-5 show that varying starting points for training, particularly within these identified boundaries (Table 4), has resulted in significant changes in forecasting power.

Table 4: M-shaped patterns of SVR-RBF (AR+ALL) forecasts, extracted from Figures 3-5

		Left boundary of "M"	Left peak	Neckline	Right peak	Right boundary of "M"
Horizon =1	Training set starts	1997Q1	1995Q3	1994Q3	1994Q1	1992Q3
	MAPE	0.0226	0.0322	0.0270	0.0319	0.0212
Horizon =2	Training set starts	1997Q1	1995Q1	1994Q3	1993Q3	1992Q3
	MAPE	0.0276	0.0375	0.0293	0.0349	0.0256
Horizon =4	Training set starts	1997Q1	1995Q1	1994Q3	1993Q1	1992Q1
	MAPE	0.0326	0.0405	0.0311	0.0389	0.0318

These empirical results suggest some interesting findings. Overall, longer training sets are more likely to produce better forecasts, despite some exceptions. Moreover, a small change in training sizes within the identified 'M' shape could lead to dramatic variations in forecast errors. Our evidence signals that the identified boundaries should be avoided as the starting point for training models, and then coupled with larger training sizes, lower errors and more consistent results can be obtained for forecasting Australia's gasoline consumption (see Figures 3-5). From a data science perspective, these patterns would be merely regarded to be 'abnormal'. However, they are also associated with some systematic influence. From an economics perspective, it is tempting to dig deeper and do more with the identified patterns. Can we characterise them and what can we learn from them? Are these systematic variations associated with or shaped by some economic events?

Figures 3-5 illustrate deteriorated forecasts for training sets starting between 1992 and 1997. Just before this period, Australia had entered its last recession, namely the early 1990s recession. According to the Reserve Bank of Australia, the 1991-92 recession mainly resulted from Australia's efforts to address excess domestic demand, to reduce inflation, and to control speculative behaviour in commercial property markets. One painful consequence is that Australia's unemployment rate reached a recorded high by the end of 1992 (over 11 percent), and it took several years to return to pre-recession levels. We plotted the pre-and post-recession relationship between Australia's inflation rate and aggregate economic activity[1] in Figure 6, which starts from 1978Q1 as the starting point of the most recent consistent data on Labour Force, Australia. After the early 1990s recession, there is a clear flattening of the Phillips curve in Australia. In addition to the role of recession, this shift may be attributed, in part, to globalisation which leads to diminishing price sensitivity to domestic demand pressure (see Kabukçuoğlu and Martínez-García 2018 for evidence from other economies).

[1] "The New Keynesian model postulates that nominal rigidities lead to the non-neutrality of monetary policy in the short run and to an exploitable trade-off between inflation and aggregate economic activity" (Kabukçuoğlu and Martínez-García, E. 2018, p.46). In the Phillips curve literature (see Rafiq 2014), economic activity is commonly measured by the quarterly unemployment rate.

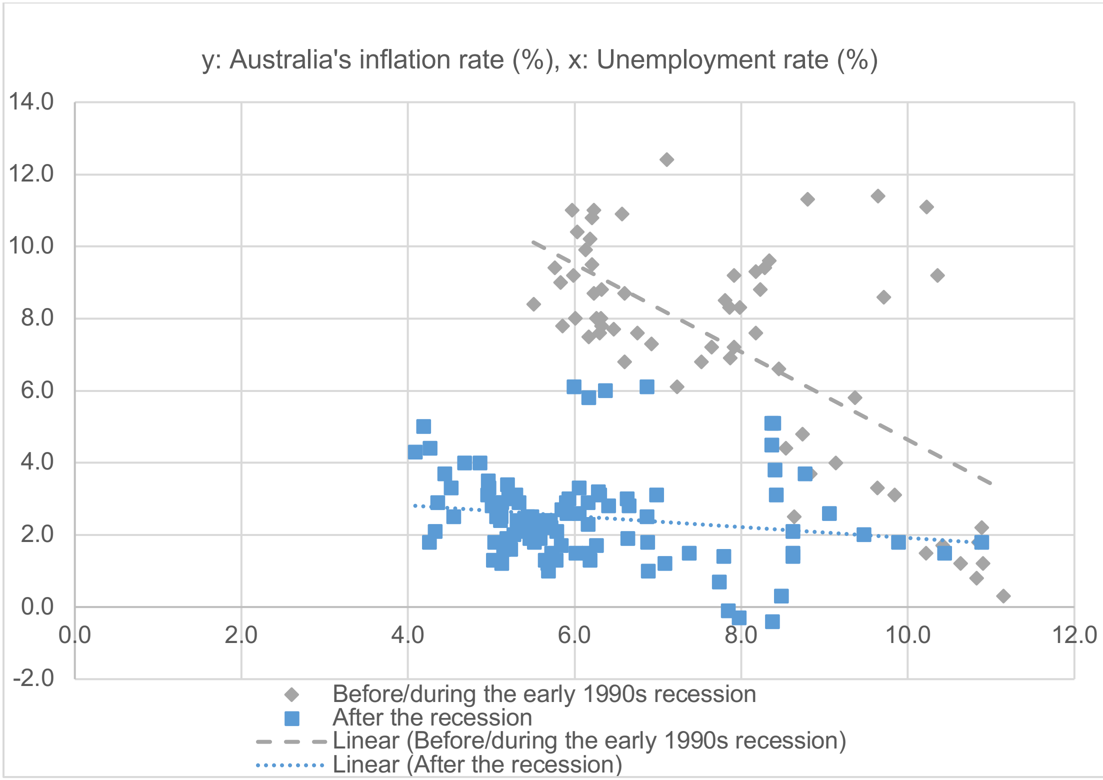


Figure 6: Flattening of the Phillips Curve in Australia

After Australia's last recession, a sub-period is rather difficult to be explained by the Phillips curve, during which there was the absence of the short-run trade-off between inflation and unemployment. From 1995Q3 (inflation=5.1% & unemployment=8.4%) to 1997Q3 (inflation=-0.4% & unemployment=8.4%), the inflation rate dropped significantly while the unemployment rate remained steady and high, and their relationship almost formed a straight line vertically (see Figure 6). This was, in part, attributed to the sharp fall in inflationary expectations after the recession. Following the 1991-92 recession, some structural changes occurred, for example, shrinking full-time employment and growing part-time/casual opportunities in the labour market, and a move to a more open economy accompanied by deregulation of the financial system and the transport sector. It takes time to completely adapt radical changes. For example, it was until 1997 that the falling inflationary expectations in the bond market fully reflected the lower trend rate of inflation (Gruen *et al.* 1999). The identified 'M' shape covers this 'weird' period with the 1991-92 recession and associated structural changes in Australia's economy. The nonlinear model is rather sensitive to changing rules of the data. For training sets starting within the 'M' shape covering 1992-1997 (i.e., a shifting/adapting period for Australia's economy), the number of observations is insufficient to capture such drastic fluctuations, which may contribute to greater forecast errors and less consistent performance of nonlinear SVR.

In light of the evidence, a certain connection between 'noise' in the data and structural changes in the economy can be established. Despite that, we can draw on studies using econometrics approaches to better inform the outputs of machine learning. For example, Mensi *et al*. (2015) found that their identified structural breaks in oil price volatility are linked to extreme economic and political events. Clark *et al*. (2020) concluded that "structural changes such as the Great Moderation or unusual periods such as the recent Great Recession can lead to significant shifts in the sizes of forecast errors and, in turn, forecast uncertainty" (p.17). Figure 6 illustrates a much clearer inflation-unemployment dynamics after the year of 1997 or before the recession. Coincidently, the models with training starting points outside of the 'M' shape tend to deliver more accurate forecasts than those falling within it, even with smaller training sizes but under more stable economic conditions (i.e., after 1997).

6.3 Experiment 3 - Selecting the optimal model specification for demand forecasting

Given our earlier discussion on the significance of training set selection, we trained the SVR-RBF models with different feature combinations using all training sets. The MAPE results for each feature combination along with the corresponding optimal training period are summarised in Table 5 (Horizon=1), Table 6 (Horizon=2) and Table 7 (Horizon=4). When Horizon=1, the model with all features, trained over the period of 1974Q3-2017Q2, produced the best hold-out-sample forecast among all model specifications (features*training sets). For Horizon=2&4, the best forecasting model is the all-feature specification trained by 1978Q1-2017Q and 1979Q1-2017Q respectively. In addition to highlighting the role of training set selection, this experiment allows us to choose the optimal model for demand forecasting. Tables 6-8 illustrate accurate hold-out-sample forecasts, with the sign of percentage errors suggesting overprediction (+)/underprediction (-).

Table 5: Hold-out-sample forecasts of SVR-RBF models with different features (Horizon=1)

Hold-out sample	Real demand	AR+Basic (1976Q3-2017Q2)		AR+Basic+HCR (1979Q1-2017Q2)		AR+Basic+POV (1974Q3-2017Q2)		AR+Basic+FCE (1976Q1-2017Q2)		**AR+ALL (1974Q3-2017Q2)**	
		Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%
17Q3	4703.10	4674.79	-0.60%	4663.87	-0.83%	-0.95%	-0.95%	-0.95%	-0.95%	-0.95%	-0.95%
17Q4	4804.60	4789.87	-0.31%	4789.29	-0.32%	-1.01%	-1.01%	-1.01%	-1.01%	-1.01%	-1.01%
18Q1	4476.80	4585.3	2.42%	4589.31	2.51%	2.25%	2.25%	2.25%	2.25%	2.25%	2.25%
18Q2	4385.30	4439.44	1.23%	4431.68	1.06%	0.68%	0.68%	0.68%	0.68%	0.68%	0.68%
18Q3	4416.10	4636.63	4.99%	4638.17	5.03%	4.09%	4.09%	4.09%	4.09%	4.09%	4.09%
18Q4	4592.40	4662.32	1.52%	4663.53	1.55%	0.82%	0.82%	0.82%	0.82%	0.82%	0.82%
19Q1	4283.10	4478.8	4.57%	4447.19	3.83%	3.91%	3.91%	3.91%	3.91%	3.91%	3.91%
19Q2	4278.80	4270.29	-0.20%	4283.72	0.11%	-0.04%	-0.04%	-0.04%	-0.04%	-0.04%	-0.04%
MAPE		0.0198 (1.98%)		0.0191 (1.91%)		0.0183 (1.83%)		0.0177 (1.77%)		0.0172 (1.72%)	

Note: Error in percentages = (Forecast - Real demand)/Real demand, +: Overprediction; -: Underprediction

Table 6: Hold-out-sample forecasts of SVR-RBF models with different features (Horizon=2)

Hold-out sample	Real demand	AR+Basic (1975Q1-2017Q2)		AR+Basic+HCR (1978Q1-2017Q2)		AR+Basic+POV (1978Q3-2017Q2)		AR+Basic+FCE (1979Q1-2017Q2)		**AR+ALL (1978Q1-2017Q2)**	
		Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%
17Q3	4703.10	4659.70	-0.92%	4647.99	-1.17%	4642.59	-1.29%	4799.21	2.04%	4648.57	-1.16%
17Q4	4804.60	4803.44	-0.02%	4799.64	-0.10%	4798.32	-0.13%	4599.27	-4.27%	4791.43	-0.27%

18Q1	4476.80	4614.98	3.09%	4610.69	2.99%	4602.04	2.80%	4472.69	-0.09%	4601.79	2.79%
18Q2	4385.30	4477.46	2.10%	4448.36	1.44%	4417.51	0.73%	4621.24	5.38%	4440.86	1.27%
18Q3	4416.10	4641.34	5.10%	4643.91	5.16%	4630.93	4.86%	4707.44	6.60%	4616.73	4.54%
18Q4	4592.40	4721.33	2.81%	4725.92	2.91%	4715.21	2.67%	4454.47	-3.00%	4643.16	1.11%
19Q1	4283.10	4463.73	4.22%	4454.59	4.00%	4464.29	4.23%	4282.15	-0.02%	4442.15	3.71%
19Q2	4278.80	4276.91	-0.04%	4282.32	0.08%	4282.7	0.09%	4652.3	8.73%	4302.52	0.55%
MAPE		0.0229 (2.29%)		0.0223 (2.23%)		0.0210 (2.10%)		0.0214 (2.14%)		0.0193 (1.93%)	

Table 7: Hold-out-sample forecasts of SVR-RBF models with different features (Horizon=4)

Hold-out sample	Real demand	AR+Basic (1984Q1-2017Q2)		AR+Basic+HCR (1979Q3-2017Q2)		AR+Basic+POV (1979Q1-2017Q2)		AR+Basic+FCE (1984Q1-2017Q2)		**AR+ALL (1979Q1-2017Q2)**	
		Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%	Forecast	Error%
17Q3	4703.10	4666.66	-0.77%	4687.86	-0.32%	4706.66	0.08%	4816.33	2.41%	4678.36	-0.53%
17Q4	4804.60	4792.95	-0.24%	4808.09	0.07%	4830.9	0.55%	4599.24	-4.27%	4808.11	0.07%
18Q1	4476.80	4583.14	2.38%	4595.97	2.66%	4608.33	2.94%	4453.86	-0.51%	4615.08	3.09%
18Q2	4385.30	4443.74	1.33%	4471.06	1.96%	4429.02	1.00%	4668.92	6.47%	4477.13	2.09%
18Q3	4416.10	4635.01	4.96%	4661.28	5.55%	4669.87	5.75%	4744.65	7.44%	4555.86	3.16%
18Q4	4592.40	4759.83	3.65%	4771.13	3.89%	4759.9	3.65%	4491.3	-2.20%	4719.76	2.77%
19Q1	4283.10	4552.98	6.30%	4482.89	4.66%	4515.24	5.42%	4302.84	0.46%	4484.93	4.71%
19Q2	4278.80	4379.49	2.35%	4318.69	0.93%	4323.61	1.05%	4679.31	9.36%	4336.92	1.36%
MAPE		0.0275 (2.75%)		0.0251 (2.51%)		0.0256 (2.56%)		0.0244 (2.44%)		0.0222 (2.22%)	

Using the optimal model, we extended the forecast over the period of 2019Q3-2020Q2 to assess its out-of-sample performance (see Appendix B). The percentage errors for 2019Q3 and 2019Q4 are 3.66% and 1.90% respectively. However, the model overvalued Australia's automobile gasoline demand by 6.59% for 2020Q1 and an even much worse overprediction for 2020Q2 when the COVID-19 pandemic took hold. During the outbreak of COVID-19 pandemic, the Australian government started to implement a series of regulations/restrictions from the middle of March 2020. After that, travel activities were significantly suppressed. The field survey conducted by Beck and Hensher (2020a) suggests a reduction of over 50% in weekly household trips during Australia's initial restrictions, in which the biggest reduction was its automobile sector. Even under easing resections after the 8^{th} of May, aggregate travel is still one-third lower than pre-COVID-19 levels (Beck and Hensher 2020b). Our forecasting error for the second quarter of 2020 (i.e., an overprediction of 38.66%) appears to be consistent with Australia's reduced travel demand (Beck and Hensher 2020a, 2020b) after the outbreak of COVID-19. Assuming a scenario without this pandemic, the forecast might be much closer to what would be consumed. Given the strong correlation between travel activities and fuel consumption, Beck and Hensher's findings provide a way to assess the external validity of our ML forecasting model under *normal and stable* situations, with its out-of-sample (pre-COVID-19) MAPE excluding 2020Q1-Q2 being 0.0278, similar to its hold-out-sample MAPE of 0.0222. It means that, with a small size sample, our model has excellent forecasting power in the absence of special events. However, under event-driven uncertainty, this type of forecasting model, established on identified characteristics in the time series, is unable to precisely anticipate the influence of such discrete events.

7. Conclusions and Future Research

With ML algorithms becoming increasingly user-friendly, a rising concern is that they might be applied naively or their results might be interpreted improperly (Mullainathan and Spiess 2017). Using Australia's automobile gasoline demand as the empirical setting, we discovered some systematic patterns, which could significantly diminish the forecasting accuracy rate. Our investigation has uncovered the important role of training set selection on ML forecasting accuracy in the case of small samples. The results show that the full training set failed to yield the best performance, and hence, we suggest that future practitioners should vary their training sizes so as to choose the optimal one. Secondly, rather than simply treating them as noise, we added in economic intuition and the implications behind the 'black-box'. An interesting and novel observation is that these patterns represent a shifting/adapting period for Australia's economy. These findings, in turn, suggest that economic events (e.g., recession) or structural changes in the economy could signal useful information on designing a better ML forecasting model. Our evidence reinforces the need for economic inputs into the machine learning environment. Lastly, through comparisons among several experiments, we got the best model delivering the lowest hold-out-sample error for four-quarter-ahead forecasting. The results of a forecast for the gasoline demand over the period of 2019Q3-2020Q2 show good prediction performance of our ML model under normal situations. However, ML in general in isolation from other informative sources, is not suitable for *rare and uncertain* events with low-probability-of-occurrence but high-impact disruptions, which is not the purpose of this current study. This indeed signals an alert that policymakers need to constantly monitor dramatic changes in the market, regularly revisit their forecasting models and update demand predictions correspondingly.

In future research, our method will be extended in several ways. On the one hand, we plan to develop a risk-exposure model to anticipate the characteristics of extreme events and to improve its response to external uncertainty, by embedding alternative machine learning techniques (e.g., Hewamalage *et al.* 2021; Kong *et al.* 2020) and promising econometrics approaches (e.g., Clark *et al.* 2020; Lyócsa *et al*. 2020). On the other hand, this typical event can be used as a natural experiment to investigate important policy levers such as telecommuting, which may be another future research avenue to monitor and understand the associated impacts on travel behaviour, energy consumption and productivity. In addition to energy security, environmental concerns in the transport sector are growing. For example, annual emissions (carbon-dioxide equivalents) from Australia's transport sector increased from 82 million tonnes in 2005 to 101 million tonnes in 2019. The best way to promote green transport and sustainable growth is to switch to vehicles using clean energy, mainly battery electric vehicles, and according to a recent Grattan Institute Report (Wood *et al*., 2021), Australia must increase its annual electric vehicle sales from 7000 to more than a million by 2035. As such, electricity demand forecasting is expected to be a major research theme, and the models used in this current study can be a starting point while calling on the development of more appearling methods.

References

ABS (2019). Survey of Motor Vehicle Use, Australia, Australian Bureau of Statistics, Australian Government.

ABS (2020). Motor Vehicle Census, Australia, Australian Bureau of Statistics, Australian Government.

ABS (2008). Survey of Motor Vehicle Use, Australia, Australian Bureau of Statistics (ABS), Australian Government.

Allen, G. (2019) AAAS: Machine learning 'causing science crisis', https://www.bbc.com/news/amp/science-environment-47267081. Accessed on September 22, 2020.

Azadeh, A., Boskabadi, A., & Pashapour, S. (2015). A unique support vector regression for improved modelling and forecasting of short-term gasoline consumption in railway systems. International Journal of Services and Operations Management, 21(2), 217.

Ali, E. S., & Abd Elazim, S. M. (2018). Mine blast algorithm for environmental economic load dispatch with valve loading effect. Neural Computing and Applications, 30(1), 261-270.

Abdelaziz, A. Y., Ali, E. S., & Abd Elazim, S. M. (2016). Implementation of flower pollination algorithm for solving economic load dispatch and combined economic emission dispatch problems in power systems. Energy, 101, 506-518.

Barutcuoglu, Z., Schapire, R. E., & Troyanskaya, O. G. (2006). Hierarchical multi-label prediction of gene function. Bioinformatics, 22(7), 830–836.

Beck, M. J., & Hensher, D. A. (2020a). Insights into the impact of COVID-19 on household travel, work, activities and shopping in Australia – the early days under restrictions. Transport Policy, 96, 76–93.

Bastardie, F., Nielsen, J. R., Andersen, B. S., & Eigaard, O. R. (2013). Integrating individual trip planning in energy efficiency–Building decision tree models for Danish fisheries. Fisheries Research, 143, 119-130.

Beck, M. J., & Hensher, D. A. (2020b). Insights into the impact of COVID-19 on household travel and activities in Australia – The early days of easing restrictions. Transport Policy. doi:10.1016/j.tranpol.2020.08.004.

Besagni, G., & Borgarello, M. (2018). The determinants of residential energy expenditure in Italy. Energy. 165, 369-386.

Beyca, O. F., Ervural, B. C., Tatoglu, E., Ozuyar, P. G., & Zaim, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. Energy Economics. 80, 937-949.

Clark, T. E., McCracken, M. W., & Mertens, E. (2020). Modeling time-varying uncertainty of multiple-horizon forecast errors. The Review of Economics and Statistics, 102(1), 17–33.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Dahl, C.A., & Sterner, T., (1991). Analysing gasoline demand elasticities: a survey. Energy Economics 13 (3), 203–210.

Ding, Y. (2018). A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting. Energy, 154, 328-336.

Ediger, V.S. & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy, 35(3), 1701-1708.

Forouzanfar, M., Doustmohammadi, A., Hasanzadeh, S., & Shakouri G, H. (2012). Transport energy demand forecast using multi-level genetic programming. Applied Energy, 91(1), 496–503.

Geem, Z.W. (2011). Transport energy demand modeling of South Korea using artificial neural network. Energy Policy, 39(8), 4644–4650.

Georges, C., & Pereira, J. (2020). Market Stability with Machine Learning Agents. Journal of Economic Dynamics and Control, 104032. doi:10.1016/j.jedc.2020.104032

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. Energy Economics, 81, 709–727.

Gruen, D., Pagan, A., & Thompson, C. (1999). The Phillips curve in Australia. Journal of Monetary Economics, 44(2), 223–258.

Ghasemi, A., Shayeghi, H., Moradzadeh, M., & Nooshyar, M. (2016). A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management. Applied Energy, 177, 40-59.

Guo, Y., Wang, J. Y., Chen, H.X., Li, G.N., Liu, J.Y., Xu, C.L., Huang, R,G., Huang, Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Applied Energy, 221, 16–27.

Haldenbilen, & Ceylan, H. (2005). Genetic algorithm approach to estimate transport energy demand in Turkey, Energy Policy, 33(1), 89-98

Hewamalage, H., Bergmeir, C. & Bandara, K. (2021). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427.

Haque, H., Chowdhury, A. K., Khan, M. N. R., & Razzak, M. A. (2021, April). Demand Analysis of Energy Consumption in a Residential Apartment using Machine Learning. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-6). IEEE.

Jiang, Y., Ma, C. Q., Yang, X. G., & Ren, Y. S. (2018). Time-varying volatility feedback of energy prices: evidence from crude oil, petroleum products, and natural gas using a TVP-SVM model. Sustainability, 10(12), 4705.

Kabukçuoğlu, A., & Martínez-García, E. (2018). Inflation as a global phenomenon—Some implications for inflation modeling and forecasting. Journal of Economic Dynamics and Control, 87, 46–73.

Khanali, M., Akram, A., Behzadi, J., Mostashari-Rad, F., Saber, Z., Chau, K. W., & Nabavi-Pelesaraei, A. (2021). Multi-objective optimization of energy use and environmental emissions for walnut production using imperialist competitive algorithm. Applied Energy, 284, 116342.

Kazemi, A., Hamed, S.G., Bagher Menhaj, M., Reza, Mehregan, M. & Neshat, N. (2010). A hierarchical artificial neural network for transport energy demand forecast: Iran case study, Neural Network World, 20(6), 761-772.

Kunwar, N., Y., Yash, K., & Kumar, R. (2013). Area-load based pricing in DSM through ANN and heuristic scheduling. Smart Grid, 4(3), 1275-1281.

Kong, L., Sun, J. & Zhang C. (2020) SDE-Net: Equipping Deep Neural Networks with Uncertainty Estimates. Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119.

Lewis, C. (1986). Industrial and Business Forecasting Methods. Butterworth Scientific.

Li, Z., Rose, J.M. & Hensher, D.A. (2010) Forecasting automobile petrol demand in Australia: An evaluation of empirical models. Transportation Research Part A: Policy and Practice, 44(1),16-38.

Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Applied Energy, 221, 386-405.

Limanond, T., Jomnonkwao, S., & Srikaew, A. (2011). Projection of future transport energy demand of Thailand. Energy Policy, 39(5), 2754–2763.

Lyócsa, Š., Molnár, P., Plíhal, T. & Širaňová, M. (2020). Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. Journal of Economic Dynamics and Control, 119, 103980.

Mashhadi, A.R., Behdad, S., (2018). Discriminant effects of consumer electronics use-phase attributes on household energy prediction. Energy Policy 118, 346–355.

Mensi, W., Hammoudeh, S. & Yoon, S.-M. (2015). Structural breaks, dynamic correlations, asymmetric volatility transmission, and hedging strategies for petroleum prices and USD exchange rate. Energy Economics, 48, 46–60.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87–106.

Mirakyan, A., Meyer-Renschhausen, M., & Koch, A. (2017). Composite forecasting approach, application for next-day electricity price forecasting. Energy Economics, 66, 228-237.

Murat Y. S. and Ceylan, H. (2006) Use of artificial neural networks for transport energy demand modelling. Energy Policy, 34(17), 3165–3172.

Nabavi-Pelesaraei, A., Abdi, R., & Rafiee, S. (2014). Applying artificial neural networks and multi-objective genetic algorithm to modeling and optimization of energy inputs and greenhouse gas emissions for peanut production. Int. J. Biosci, 4(7), 170-183.

Nabavi-Pelesaraei, A., Fatehi, F., & Mahmoudi, A. (2014). Prediction of yield and economic indices for tangerine production using artificial neural networks based on energy consumption. International Journal of Agronomy and Agricultural Research, 4(5), 57-64.

Nabavi-Pelesaraei, A., Bayat, R., Hosseinzadeh-Bandbafha, H., Afrasyabi, H., & Chau, K. W. (2017). Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management-A case study in Tehran Metropolis of Iran. Journal of Cleaner Production, 148, 427-440.

Ou, S., He, X., Ji, W., Chen, W., Sui, L., Gan, Y., Lu, Z.F., Lin, Z.H., Deng, S.L. , Przesmitzki, S, & Bouchard, J. (2020). Machine learning model to project the impact of COVID-19 on US motor gasoline demand. Nature Energy, 5(9), 666-673.

Olanrewaju, A O., Jimoh, A A., & Kholopane, P. A. (2013). Assessing the energy potential in the South African industry: A combined IDA-ANN-DEA (Index Decomposition

Analysis-Artificial Neural Network-Data Envelopment Analysis) model. Energy, 63, 225-232.

Plakandaras, V., Gogas, P., Papadimitriou, T., & Gupta, R. (2017). The Informational Content of the Term Spread in Forecasting the US Inflation Rate: A Nonlinear Approach. Journal of Forecasting, 36(2), 109–121.

Panapakidis, I. P., & Dagoumas, A. S. (2017). Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. Energy, 118, 231-245.

Peng, L., Liu, S., Liu, R., & Wang, L. (2018). Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy, 162, 1301-1314.

Papadimitriou, T., Gogas, P., & Stathakis, E. (2014). Forecasting energy markets using support vector machines. Energy Economics, 44, 135-142.

Rafiq, S. (2014). What Do Energy Prices Tell Us About UK Inflation? Economica, 81(322), 293–310.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.

Teng, G., Xiao, J., He, Y., Zheng, T., & He, C. (2017). Use of group method of data handling for transport energy demand modeling. Energy Science & Engineering, 5(5), 302–317.

Vapnik, V. (1998). Statistical Learning Theory. John Wiley & Sons, New York.

Vapnik, V., Boser, B., & Guyon, I. (1992). A training algorithm for optimal margin classifiers. In: Fifth Annual Workshop on Computational Learning Theory. ACM, Pittsburgh, pp. 144–152.

Xiao, J., Li, Y., Xie, L., Liu, D., Huang, J. (2018). A hybrid model based on selective ensemble for energy consumption forecasting in China. Energy, 159, 534–546.

Yang, J. H., Wu, C. Y., & Zhang, G. M. (2014). Coal demand prediction in Shandong Province based on artificial firefly wavelet neural network. In Advanced Materials Research (Vol. 962, pp. 1931-1935). Trans Tech Publications Ltd.

Yu, S., Wei, Y.-M., & Wang, K. (2012). A PSO–GA optimal model to estimate primary energy demand of China. Energy Policy, 42, 329–340.

Yu, L., Zhao, Y., & Tang, L. (2017). Ensemble forecasting for complex time series using sparse representation and neural networks. Journal of Forecasting, 36(2), 122-138.

Zhang, J. L., Zhang, Y. J., & Zhang, L. (2015). A novel hybrid method for crude oil price forecasting. Energy Economics, 49, 649-659.

Zhu, B., Shi, X., Chevallier, J., Wang, P., & Wei, Y. M. (2016). An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. Journal of Forecasting, 35(7), 633-651.

Zhao, X., & Nie, X. K. (2020). Prediction error and forecasting interval analysis of decision trees with an application in renewable energy supply forecasting. Complexity, 2020, 3567894.

Appendix A: Best/worst hold-out-sample predictions under SVR-RBF with all features

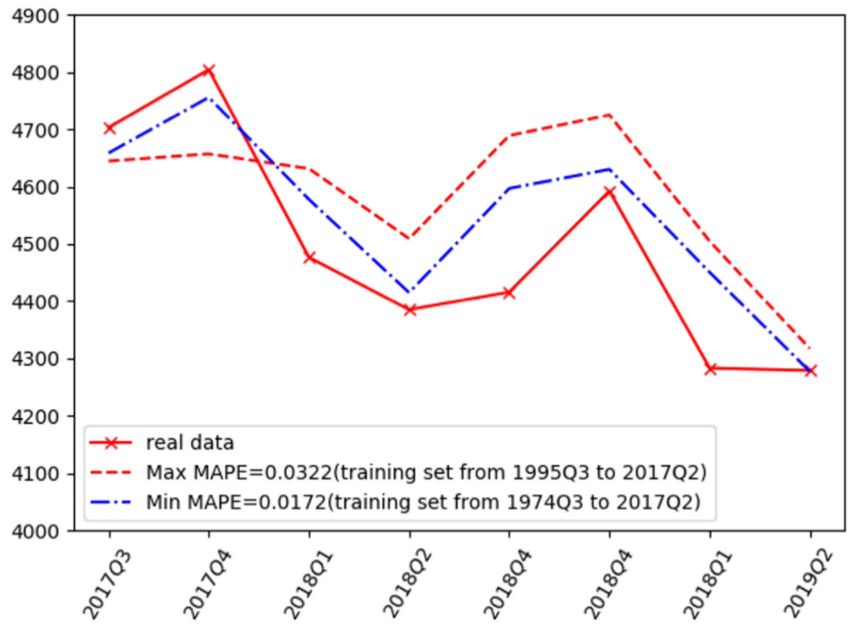


Figure A1: Horizon=1

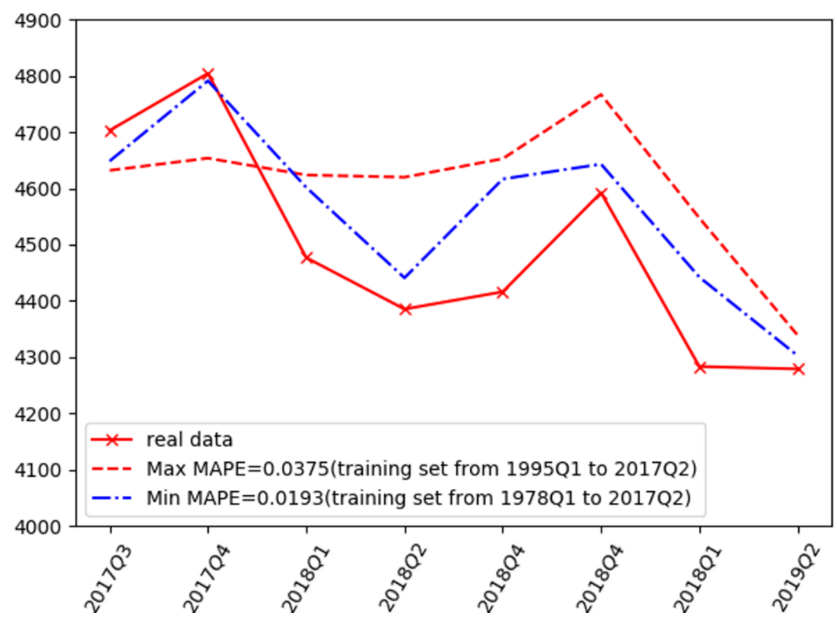


Figure A2: Horizon=2

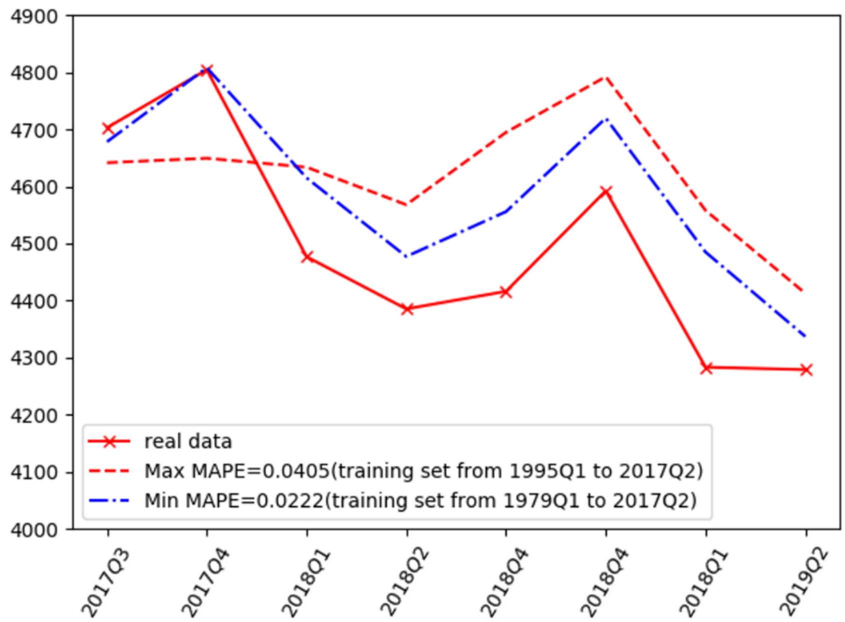
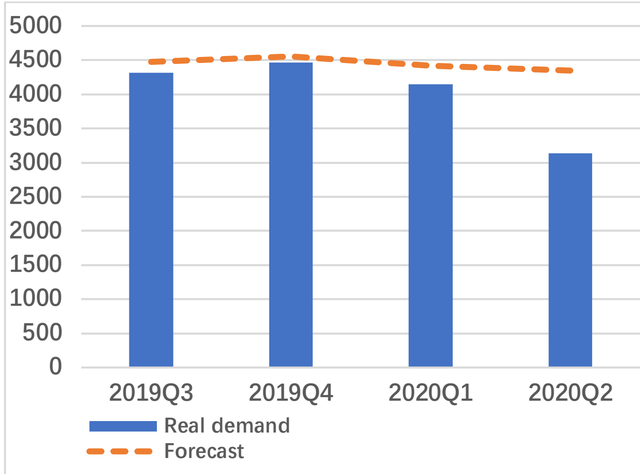


Figure A3: Horizon=4

Appendix B: Out-of-sample prediction using the optimal model for Horizon=4



Out-of-sample	Real demand	Forecast	Error	Error%
19Q3	4317.30	4475.33	158.03	3.66%
19Q4	4469.30	4554.43	85.13	1.90%
20Q1	4148.90	4422.55	273.65	6.60%
20Q2	3135.30	4347.45	1212.15	38.66%
MAPE		0.127		
MAPE excluding 2020Q2		0.0405		
MAPE excluding 2020Q1-Q2		0.0278		

Figure B: Out-of-sample prediction of SVR-RBF with all features and training set being 1979Q1-2017Q2

Appendix C:

Table C: The list of abbreviations

Abbreviation	Full Name
ABS	Australian Bureau of Statistics
ML	machine learning
COVID-19	Corona Virus Disease 2019
LPG	Liquefied Petroleum Gas
GA	genetic algorithm
ANN	artificial neural networks
MLGP	multi-level genetic programming
GMDH	group method of data handling
SVM	support vector machine
SVR	support vector regression
MLR	Multiple linear regression
GDP	Gross domestic product
GNP	Gross national product
LASSO	least absolute shrinkage and selection operator
OLS	ordinary least square

SRM	structural risk minimisation
TAGC	total automobile gasoline consumption
RGP	real gasoline price
RHGDI	real household gross disposable income
POP	population in millions
HCR	expenditures on hotels, cafes and restaurants
FCE	final consumption expenditure
POV	purchase of vehicles
AR	autoregressive
MAPE	mean absolute percentage error
MSE	mean square error
MAE	mean absolute error
RBF	Radial Basis Function
ALL	AR+All features
Basic features	national income, population and gasoline price
AR	the autoregressive component