Combination Forecasting of Energy Demand in the UK

Marco Barassi *
Yuqian Zhao †

Abstract

In more deregulated markets such as the UK, demand forecasting is vital for the electric industry as it is used to set electricity generation and purchasing, establishing electricity prices, load switching and demand response. In this paper we produce improved short-term forecasts of the demand for energy produced from five different sources in the UK averaging from a set of 6 univariate and multivariate models. The forecasts are averaged using six different weighting functions including Simple Model Averaging (SMA), Granger-Ramanathan Model Averaging (GRMA), Bayesian Model Averaging (BMA), Smoothing Akaike (SAIC), Mallows Weights (MMA) and Jackknife (JMA). Our results show that model averaging gives always a lower Mean Square Forecast Error (MSFE) than the best/optimal models within each class however selected. For example, for Coal, Wind and Hydro generated Electricity forecasts generated with model averaging, we report a MSFE about 12% lower than that obtained using the best selected individual models. Among these, the best individual forecasting models are the Non-Linear Artificial Neural Networks and the Vector Autoregression and that models selected by the Jackknife have often superior performance. However, MMA averaged forecasts almost always beat the predictions obtained from any of the individual models however selected, and those generated by other model averaging techniques.

Keywords: Demand for Energy; Forecasting; Model Averaging;
JEL Classification: C53, C55, Q47

The authors wish to thank Aurelio Antani, David Broadstock, Lutz Kilian, Charles Mason and Nicola Spagnolo for useful comments and suggestions.

*Department of Economics, University of Birmingham, Birmingham, B15 2TT, U.K.; Email: m.r.barassi@bham.ac.uk
†Department of statistics and actuarial science, University of Waterloo, ON, Canada; Email:yuqian.zhao@uwaterloo.ca

1 Introduction

Accurate and rigorous electricity demand modelling and forecasting is extremely important for energy suppliers, Independent System Operators (ISOs), financial institutions, and other participants in electric energy generation, transmission, distribution, and markets. Forecasting the demand for energy is crucial for planning periodical operations and facility expansion in the electricity sector at the various levels. For example, models for electric power load forecasting are essential to the operation and planning of a utility company. At this level, load/demand forecasting would help an electric utility to make important decisions including purchasing and generating electric power, load switching, and infrastructure development. On a broader and different level, models and forecasts of a country's energy demand may provide useful information for the implementation of specifically targeted energy policies. However, obtaining an appropriate forecasting model for electricity networks is far from being an easy task. In fact, although many modelling and forecasting methods have been developed, none can be generalized for all demand patterns. This becomes an even more pressing issue in deregulated markets such as that of the United Kingdom (UK hereafter) where demand patterns have become even more complex. Depending on the available data, their frequency, the desired nature and detail of forecasting, methodologies to obtain forecasts of the demand for energy can be broadly classified into short-term models, which use traditional time series, similar day/machine learning approaches and long-term models, which include, end-use, structural econometric models and time series models with lower frequency data.

In this paper we aim to contribute to the literature on energy demand forecasting by detailing a pseudo out-of-sample combination forecast design which uses high frequency time series data to obtain improved short-term forecast (up to 1 day) of the demand for energy produced from five different sources in the UK, averaging from a set of 6 univariate and multivariate models. With the growing deregulation of the energy industries, obtaining more accurate forecasts, has gained increasing appeal. The reason is that, since supply and demand experience wider fluctuations and energy prices may increase by a factor of ten or more during peak situations, correct and well timed demand forecasting has become vitally important for utilities. Short-term demand forecasts help to estimate load flows and to make decisions that can prevent overloading. Timely implementations of such decisions would lead to the improvement of network reliability and to the reduced occurrences of equipment failures and blackouts. On the other hand, demand forecasting is also important for contract evaluations and evaluations of various sophisticated financial products on energy pricing offered by the market.

Short-term forecasting models, which are usually from one hour to one week, use historical data and play a very important role in the operation of power systems' operating functions such energy transaction, unit commitment, security analysis, fuel scheduling and load switching. The techniques most commonly used include:

- Univariate time series models (Box-Jenkins ARIMA), Holt-Winters exponential smoothing, time series regressions and multivariate time series models such as Vector autoregressions (VARs), Bayesian VARs (BVAR) and Factor Augmented VARs (FAVAR)
- Similar day and machine learning approaches including Artificial Neural Networks (ANN),

Non-linear Autoregressive Neural Networks (NLANN), Fuzzy Logic, Support Vector Machines.

Long-term forecasting models play an important role in policy formulation and supply capacity expansion (incorporate consumer behaviour and characteristics, technology, etc). They include:

- Time series methods with Lower frequency data. These include both univariate time series models such as the traditional Box-Jenkins ARIMA, but also Seasonal ARIMA and models which include fractionally integration such as ARFIMA and also the above mentioned multivariate time series models (VAR, BVAR, FAVAR) also including cointegrated vector error correction models (VECM).
- End-use methods (electricity demand is derived from users' demand for individual requirements):
- Non-intrusive Load Monitoring Models
- Structural econometric models (Seek to establish the relationship between energy consumption and the factors that influence it) include:
- Conditional Demand Analysis (Multivariate Regressions, Stochastic Markov Chains)

Earlier attempts of short-term forecasting include Hagan and Behr (1987), Fan and McDonald (1994), Amjady (2001) and Nogales et al. (2002), who used time series regressions including temperature, humidity and past energy consumption to obtain demand projections. More recently, Filik et al. (2011) predicted yearly, weekly and hourly electric energy demand through a three stage model, showing how short-term predictions are usually more accurate and of immediate application than medium and long-term ones.

Within the literature on the UK energy system, we find several studies mostly focussed on long-term modelling and forecasting. Hunt et al. (2003) applied time series models to forecast the UK energy demand on a sectoral basis. However, pure time series models have often been criticised for not considering other important macroeconomic variables as predictors of long-run energy demand. To overcome this issue, Haas and Schipper (1998) use price elasticities, income elasticities and technical efficiency as explanatory variables to forecast energy demand in the UK and other OECD countries. Similarly, in order to forecast oil, gas, coal and total energy demand in the UK and Germany, McAvinchey and Yannopoulos (2003) constructed an econometric model incorporating variables including the price of electricity and technological progress. Cointegration models have also been used to model or forecast energy demand in the long-run. Among them, Fouquet et al. (1997) uses a cointegrated VAR to investigate the long run relationship among fuel demand, the real price level and economic activity and Sadorsky (2009) adopted panel cointegration techniques to model the long run relationship among GDP per-capita, Co_2 per-capita and the demand for renewable energy in the G7 countries.

Previous studies on energy demand forecasting can be also classified on the basis of their parameterisation, that is, whether they employed univariate and/or multivariate (time series) models. The most widely used univariate forecasting set of models is the ARMA, which can be extended to

ARIMA to account for non-stationarity, SARMA to consider seasonality and ARFIMA to consider fractional integration. Ediger et al. (2006) and Ediger and Akar (2007) employed ARIMA and SARIMA to forecast fossil fuel in Turkey, using goodness-of-fit and information criteria to select the best forecasting models. However, Sumer et al. (2009) focussed on the importance of capturing the seasonal effects contained in energy demand and used ARIMA, SARIMA and various regression models, and found that the regression model with seasonal latent variables provides more accurate forecast than the class of ARMA models.

In the earlier literature, exponential smoothing also showed reasonable forecasting ability. Badri et al. (1997) adopted several time-series models including exponential smoothing to forecast electricity peak-load in the UAE. As an extension of the simple exponential smoothing technique, Hong (2013) claimed that Holt-Winter smoothing method shows better performance. More recently, with the introduction of Artificial Neural Networks (ANN) and its application in forecasting, many studies have used ANN-type of models to forecast energy demand, also considering a number of input variables, such as macroeconomic and environmental variables (Chow and Leung (1996), Markham and Rakes (1998), Sözen et al. (2005), Ermis et al. (2007)). However, Maia et al. (2006) showed that a hybrid model of ARMA and ANN model shows better prediction ability of energy forecast. Similar models have also have been applied by Pao (2006) and Kurban and Filik (2009), who used Non-linear Autoregressive Neural Networks –(NARNN) to forecast electricity demand in Taiwan and Turkey, respectively. Geem and Roper (2009) employed the NARNN model to forecast energy demand for South Korea, and they indicated that NARNN produce more accurate predictions than linear regressions and exponential smoothing.

Among multivariate time series models, vector autoregressions (VAR) are the probably the most popular for forecasting purposes. García-Ascanio and Maté (2010) used a VAR to forecast electric power demand, but they argued that VARs just shows poor predictive ability compared with more advanced multivariate models. Bayesian VARs –BVAR have often been found to improve the forecast accuracy of the basic VAR also overcoming the problem of over-fitting. Crompton and Wu (2005) forecast coal, oil, gas and hydro energy demand in China through a BVAR; Francis et al. (2007) used the BVAR to study the relationship between real gross domestic product per capita and energy demand in Caribbean countries, and obtained forecasts for energy demand in those countries. Recently, VAR models have been further developed so to include factors extracted from a set of potentially numerous predictors so to obtain a factor augmented vector autoregressive model –FAVAR (see (Chudik and Pesaran, 2011) among others). Baumeister et al. (2016) used a FAVAR model to forecast gasoline price for the US, showing that their predictions are more accurate than those of standard VARs.

Clearly, there is a wide variety of models and parameterisations among which to choose for the purpose of obtaining energy demand forecasts, and thus, the choice of which criterion to use in order to select the optimal forecasting model becomes another important issue (see Pao (2006) and García-Ascanio and Maté (2010) for comparative studies). Lai et al. (2008) used the mean squared error –MSE, the mean absolute percentage error –MAPE and the mean squared percentage error –MSPE to compare the accuracy of rival forecasting models. Moreover, models can be selected by minimisation of information criteria obtained in-sample for each individual estimated model (i.e. Akaike Information Criterion - AIC, and Bayesian Information Criterion - BIC). Within this stream of literature, Hansen (2007) and Hansen (2008) showed that Mallows' information criteria may often select models that providing more precise forecasts, and later Hansen and Racine (2012) and Hansen (2014) proposed a cross-validation criteria for model selection based on the Jackknife.

Since the seminal article of Bates and Granger (1969) model averaging has been widely used to produce more accurate forecasts than the individual optimal models. Hendry and Clements (2004) and Timmermann (2006) showed that simple combinations often give better performance than more sophisticated approaches. Further, using a frequentist approach, Granger and Ramanathan (1984) proposed the use of coefficient regression methods to determine the magnitude of the weights of individual models in the averaging process. However, the most popular average method is Bayesian model averaging (Madigan and Raftery, 1994), where the averaging weights are calculated based on empirical data and uses the Bayesian Information Criterion of the individual models as weight in the averaging function. Anderson and Burnham (2002), suggested to use AIC criteria to replace BIC criteria in model averaging, while Hansen (2007) and Hansen (2008) have continued this line of research showing that model combination based on Mallows' criterion, asymptotically leads to forecasts with the smallest possible mean squared error. Guidolin and Timmermann (2007) proposed a different time varying weight combination scheme where weights have regime switching dynamics. More recently, Hansen and Racine (2012) proposed a "jackknife model averaging" (JMA) estimator which selects the weights by minimizing a cross-validation criterion showing that their method is asymptotically optimal.

In this paper, in order to overcome some of the methodological issues above and produce improved forecasts of the demand for energy in the UK, we use a forecasting approach based on model averaging of several popular linear or non-linear, univariate and multivariate forecast models. Specifically, we first obtain the forecasts from sets of ARMA, Holt-Winters Smoothing (HWS), Non Linear Autoregressive Neural Networks (NLANN), Vector Autoregressions (VAR), Bayesian VAR (BVAR), and Factor Augmented VAR (FAVAR) models. Forecasts are generated from each of these sets of models and within each set they are then compared and selected using rankings based on four different information criteria (AIC, BIC, Mallows' and Jackknife). The best models as selected by each of the different information criteria within the individual model set and are averaged using six different combination weight metrics including Simple Model Averaging (SMA), Granger-Ramanathan Model Averaging (GRMA), Bayesian Model Averaging (BMA), Smoothing Akaike (SAIC), Mallows Weights (MMA) and Jackknife (JMA).

The paper is structured as follows: the next section presents the data and explains the pre-analysis necessary to transform the raw series into workable variables. Section three outlines the individual forecast models used and the metrics used for selection and weighting purposes. Comparisons of the performance of individual models, forecasting methods, information criteria, averaging functions are discussed in section four. Section five reports the forecasts of the levels demand for energy in the UK, a summary concludes.

2 Data and Pre-analysis

2.1 Data

We collected 30 minutes data from National Gridwatch website [1] ranging from 00:00:00 of the 21 December 2013 to 00:00:00 of the 21 March 2016, for a total of 39287 observations for each energy demand process. The data relate to energy demand for the entire UK plus import, minus export less un-metered sources. Here, we shall assume that supply matches demand at all times. The UK gridwatch provides energy demand data disaggregated by different types of energy sources, and in this paper, we will concentrate on the most important five types of sources including coal, nuclear, combined cycle gas turbines (CCGT hereafter), wind and hydro-power.

Note that, energy providers in the UK differ quite substantially between them for the choice of fuel sources used to provide energy as can be seen from the table below which list most of the utilities in the UK.

Source: `http://electricityinfo.org/fuel-mix-of-uk-domestic-electricity-suppliers/`

Note that, beside the strong deregulation, given the differences in fuel mix used by energy suppliers in the UK, accurate short-term forecasting of the demand for energy produced from the different sources becomes even more appealing as it has an important impact on the prices charged, the choice of production and not least general demand management and load switching.

In this paper, using the data described above, we obtain 30-minutes to one-day predictions of the demand for energy as produced from these five energy sources, that is we obtain forecasts for the 22 March 2016, at the following times 00:30, 01:00, 02:00, 04:00, 06:00, 08:00, 16:00 and 24:00.

Insert Figure [2] about Here

Figure 2 plots the levels of the demand for energy generated from each of the five sources. It is noticeable how demand for CCGT and wind sourced energy are relatively more volatile, while nuclear is more steady. The main reason is that CCGT, though is an efficient way to use gas and turbines are fast to get online, they use relatively expensive fuel. Thus, these cycling plants will ramp up and down during the day, and are usually used more during peak hours. As for the wind turbines, they are expensive but wind is cheap, however, the strength of wind is not constant and it varies from zero to storm force. This means that wind turbines cannot not produce the same amount of electricity all the time and there are be times when they produce no electricity at all. On the other hand, once on, nuclear power stations run flat out, the cost of fuel is insignificant and this explains the much lower volatility in its demand.

Still, regardless of the source, there exist strong“seasonal” components in the data. These periodicities vary in frequency as seasons, ”day-of-the-week” and ”hour-of-the-day”. Therefore, it is necessary to remove these periodic effects prior to analyse the data and obtain the forecasts for each model of the model sets.

[1] `http://www.gridwatch.templar.co.uk/`

Table 1: Fuel Mix of UK Energy Suppliers

Supplier	Coal	Gas	Nuclear	Renewable	Other	CO_2	Nuclear Waste
British Gas	2	30	34	33	0.7	0.137	0.0024
Bulb	0	0	0	100	0	0	0
Co-operative Energy	36.6	34.2	13.4	9.8	6	0.493	0.00094
E.On	18.7	32.4	12.8	29	7.1	0.328	0.001
Ecotricity	0	0	0	100	0	0	0
EDF Energy	14.5	8.6	64.3	12.3	0.3	0.167	0.0045
Extra Energy	19	33	13	28	7	0.339	0.0009
First:Utility	18.9	32.7	12.9	28.3	7.2	0.33	0.0009
Flow Energy	18.9	32.7	12.9	28.3	7.2	0.331	0.0009
GnERGY	34	25.6	21.6	16.7	2.1	0.418	0.00151
Good Energy	0	0	0	100	0	0	0
Green Energy UK	0	70.6	0	29.4	0	0.134	0
Green Star Energy	0.1	0.1	0	99.8	0	0.002	0.00001
iSupplyEnergy	38.7	36.2	14.2	4.6	6.3	0.528	0.001
LoCO2 Energy	0	0	0	100	0	0	0
Npower/RWE	16	66	1	16	1	0.408	0.00008
Octopus Energy	1	1	1	97	0	0.013	0.00004
OVO Energy	0	46.9	0	53.1	0	0.183	0
Scottish Power	34	36	3	26	1	0.46	0.0002
So Energy	18.9	32.7	12.9	28.3	7.2	0.331	0.0009
Spark Energy	46.8	27.1	8.4	11.9	5.8	0.579	0.0007
SSE	25	35	7	29	4	0.38	0.00047
Utilita	19	33	13	28	7	0.332	0.00091
UK Average	17	32.3	23.7	24.3	2.5	0.29	0.0017

2.2 Removing the Deterministic Components

As highlighted above, energy demand data always contains significant periodic patterns in both the short and long-term. In this subsection, we describe the three methods that we shall use to remove these periodic patterns in turn. In what follows we will first fit an up to 5th degree Chebyshev polynomial to remove the season of the year component, a moving average based technique to remove the day of the week effect and finally a 48th order differencing to eliminate a day/night peak/off-peak hour effect.

Cuestas and Gil-Alana (2016), recently showed the ability of Chebyshev polynomials to fit long-term cyclical patterns. Chebyshev polynomials are based on orthogonal cosine functions of time, such that a linear combination of these functions can flexibly approximate most cyclical patterns. The higher the order of the polynomial, the more non-linear is the cyclical pattern that can be approximated. Following Cuestas and Gil-Alana (2016) we define the polynomial as,

$$P_{i,n}(t) = \sqrt{2}cos(i\pi(t-0.5)/n), \quad t = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, \tag{1}$$

where i is the order of Chebyshev polynomial. Specifically, when i equals to 0, it gives a linear constant function with $P_{0,n}(t) = 1$. Since any empirical process y_t can be decomposed between a deterministic and a stochastic part, and if the deterministic term approximated by Chebyshev polynomials, then we can have,

$$y_t = \sum_{i=0}^{m} \theta_i P_{i,n}(t) + x_t, \quad t = 1, 2, 3, \ldots, \tag{2}$$

where x_t is assumed to be the stochastic part of the model, and the order of Chebyshev polynomials is determined by the significance of parameters θ_i. The parameter θ_i can be estimated by,

$$\hat{\theta} = (\sum_{t=1}^{n} P_n(t)P_n(t)')^{-1}(\sum_{t=1}^{n} P_n(t)y_t), \tag{3}$$

Finally the de-seasonalised process y_t^* is,

$$y_t^* = y_t - \sum_{i=0}^{m} \hat{\theta}_i P_{i,n}(t), \tag{4}$$

Because our data set span over 2 years and 3 months, it should contain two complete season cycles in the demand for each energy source. Figure 3 illustrates the seasonal patterns removed from the demand of energy obtained from the five sources. From these plots, we can see that all energy sources reach their peaks during winter period and fall down during summer except for CCGT. Once the seasonal long-term cyclical pattern has been removed, it is necessary to remove shorter term periodic patterns.

Insert Figure [3] about here

In the short-run, there are two more periodic elements which need to be filtered out: week-day, and peak/off-peak (or day-night) effects. The week-day element can be removed by adopting a specific moving average method. Since the data frequency is 30 minutes, following Weron (2007), we set the moving average length l equal to 336, which are the number of half-hours in one week. To obtain the moving average of each point in the series, we need to make sure that data in front and behind any given data point which we obtain, is of the same length. Hence, we consider $l+1 = 337$ observations for each moving average window. The moving average component is calculated through,

$$\hat{m}_t = \frac{1}{l+1}(\sum_{i=1}^{l/2} y_{t-i}^* + y_t^* + \sum_{i=1}^{l/2} y_{t+i}^*), \tag{5}$$

where m_t is the moving average term and y_t^* is the de-seasonalised series obtained in the first step. Hence, the deviation from the moving average is given as,

$$w_k = y_{k+l,j} - \hat{m}_{k+l,j}, \tag{6}$$

where k is the index in one moving average window, and j is the number of moving average windows. Thus, the day-of-the-week cyclical component can be obtained as,

$$\hat{s}_k = w_k - \frac{1}{l}\sum_{i=1}^{l} w_i, \tag{7}$$

Therefore, the day-of-the-week filtered data is defined as,

$$y^*_{t,l} = y^*_t - \hat{s}_t, \tag{8}$$

Figure 4 illustrates the fitted weekly component of the demand for energy obtained for each source. From the plots, it is apparent that some energy sources have a more pronounced day-of-the-week effect compared to others. Specifically, the demand for coal, nuclear and CCGT produced energy are relatively higher in weekdays as compared with weekend. This pattern, instead, does not appear so clearly in the demand for energy produced from wind and hydro sources. This is probably because the first three energy are the main sources of electricity used in industrial processes and businesses, and also, the other two are more strongly dependent on weather conditions and cannot necessarily provide energy in a steady manner. Apart from the week-day effect, it is reasonable to expect that a peak/off-peak or day/night effect exists too in the energy demand series and this will also be dealt with in what follows.

Insert Figure [4] about Here

After removing the season-of-year and week-day components, we remove the peak/off-peak and day/night effect by taking the $s^{th} = 48$ order difference of each of the demand for energy of each ω source:

$$y^*_{\omega,t,s} = y^*_{\omega,t,l} - y^*_{\omega,t-s,l}, \tag{9}$$

After this preliminary analysis, we move forward to the second step of our forecast procedure where the obtained (now purely stochastic) energy demand series $y^*_{\omega,t,s}$ for each source ω will be modelled and forecast by univariate and multivariate time series as well as neural networks models, as briefly described in next section.

Insert Figure [5] about Here

3 Methodology

In this section, we will briefly outline the methodology that we shall use to obtain the forecasts of the demand for all the energy sources. Firstly, we will describe the six types of forecast models going from univariate to multivariate ones; secondly, we will present the four types of criteria used to select optimal forecasting model within each of the model sets (or)classes). Lastly, we discuss the model averaging techniques and show they are used to produce improved forecasts.

3.1 Forecast Models

In this subsection, we will introduce the six classes of models which we use to fit and forecast the filtered series. The univariate models are autoregressive moving average (ARMA), Holt Winter Smoothing (HWS), Non-linear Autoregressive Neural Network (NARNN) model, and the multivariate models are Vector autoregression (VAR), Bayesian VAR and Factor Augmented VAR models. Recall that ω denotes the type of fuel used as a source of energy.

1. **ARMA**
The first forecasting model is the traditional stationary ARMA model. It is the most commonly used forecast model also in the energy field and usually constitutes a benchmark against which other forecasting techniques are compared. Among others, Ediger et al. (2006) and Ediger and Akar (2007) used ARIMA models to forecast Turkish fossil fuel demand. Also, Sumer et al. (2009) employed the ARMA class of models to predict electricity demand. In this paper, since the periodic components and non-stationarity have been removed, we use the following simple ARMA(p,q) parameterisation:

$$y_{\omega,t+h-1} = \alpha + \sum_{i=1}^{p} \beta_i y_{\omega,t-i} + \sum_{j=1}^{q} \gamma_j \epsilon_{\omega,t-j} + \epsilon_{\omega,t+h-1}, \tag{10}$$

where h is the period forecast ahead. The model is estimated by ordinary least squares and the optimal lag length p and q are determined by information criteria. Once the model has been estimated, we can use it to predict

$$f_{\omega,t+h-1} = \hat{\alpha} + \sum_{i=1}^{p} \hat{\beta}_i y_{\omega,t-i} + \sum_{j=1}^{q} \hat{\gamma}_j \epsilon_{\omega,t-j}, \tag{11}$$

The optimal forecasting model is selected from the set of ARMA(p,q) with p and $q = 1, \ldots, 12$ giving a total of 144 models estimated for each energy source.

2. **Holt-Winter Smoothing (HWS)**
The HWS belongs to the class of exponentially weighted moving average methods. The model fits the target process using its past smoothed values and gives more weight to the most recent ones such that it can be expressed as,

$$s_{\omega,t} = \alpha a_{\omega,t} + (1-\alpha)(s_{\omega,t-1} + b_{\omega,t-1}), \tag{12}$$

such that at time t, the actual value of the process is denotes by $a_{\omega,t}$, the smoothed estimate is denoted by $s_{\omega,t}$ and $b_{\omega,t}$ is the trend. In turn, the trend is formulated as,

$$b_{\omega,t} = \beta(s_{\omega,t} - s_{\omega,t-1}) + (1-\beta)b_{\omega,t-1}, \tag{13}$$

where the parameter β is the trend smoothing parameter. Therefore, the predicted value is obtained from,

$$f_{\omega,t} = s_{\omega,t} + i b_{\omega,t}, \tag{14}$$

Here we select the smoothing parameter α from $[0.7, 0.8, 0.9]$ while β is selected from $[0.1, 0.2, 0.3]$ (see Hong (2013)). Hence, we consider 9 types of HWS models.

3. **Non-linear Auto-Regressive Neural Networks** (NARNN)
The restrictions imposed by linear forecast models such as ARMA and HWS, are usually overcome by adopting more general non-linear forecast models (De Gooijer and Kumar, 1992). Artificial neural networks are a commonly used type of such non-linear models (see Sözen et al. (2005), Pao (2006) and Kurban and Filik (2009), amongst others), and have been widely used for the purpose of univariate series forecasting. Although it suffers from the criticism of not-so-much underlying economic foundation, the NARNN model (Chow and Leung (1996) and Markham and Rakes (1998), amongst others) often provides better forecasting accuracy because it is able to approximate plenty of functions (Zhang, 2003).

In brief, the NARNN model is a dynamic neural network model which is built on a linear autoregressive model with feedbacks on several layers. The model regresses current dependent output signal on previous output signals, so that the model equation is defined as follows:

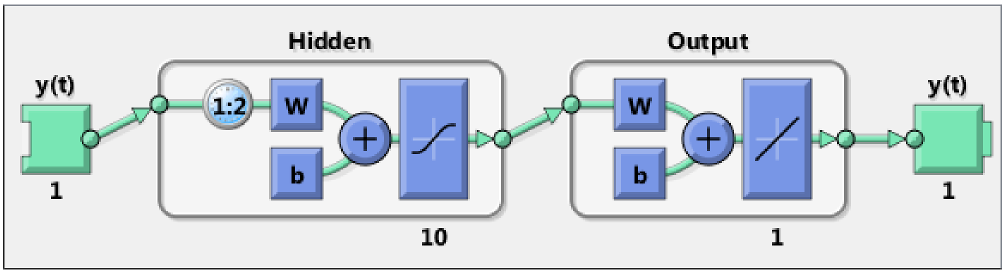
$$y_{\omega,t+h-1} = f(y_{\omega,t-1} + y_{\omega,t-2} + \cdots + y_{\omega,t-p}), \tag{15}$$

where f is a non-linear function, and p is the earliest value of signals considered. Once the model has completed training and validation, it can be used to forecast in the same fashion:

$$\hat{y}_{\omega,t+h} = \hat{f}(y_{\omega,t} + y_{\omega,t-1} + \cdots + y_{\omega,t-p+1}), \tag{16}$$

An example of the architecture of the NARNN model is showed in Figure 1. In our case, we set to 10 the number of neurons in the hidden part, and apply a back-propagation method for training as in Geem and Roper (2009).

Figure 1: **The Architecture of a NARNN model**



The source: http://uk.mathworks.com/help/nnet/ref/narnet.html

The lag length considered in NARNN ranges from 1 to 12, so that there are 12 models for each energy source.

4. **Vector Autoregression** VAR
The models introduced above are self-forecast models, where the predicted value is mainly based on the serial correlation of historical data. From the fourth model onward, we forecast the energy demand processes according to causal relationships. In these models, the energy demand processes are modelled as a system, such that the predictions for each process would be obtained from the entire system. The first causal forecast model is the standard VAR. García-Ascanio and Maté (2010) used a VAR model to forecast electric power demand in Spain.

A system of $\mathbf{y}_t$ composed by endogenous variables $y_{1,t}$, $y_{2,t}$, ..., $y_{k,t}$, where k refers to the demand for energy from each fuel source such that here $k = 5$. Thus, the VAR model with lag length p can be formulated as,

$$\mathbf{y}_{t+h-1} = \sum_{i=1}^{p} \Phi_i \mathbf{y}_{t-i} + \epsilon_t, \tag{17}$$

where each Φ_i is a $K \times K$ coefficient matrix, and ϵ_t is a K-dimensional vector of errors terms with mean vector zero and diagonal variance covariance matrix Σ. We estimate each VAR by maximum likelihood (MLE). Then, we use the in-sample estimation results and iterate forward to obtain the out-of-sample predictions.

$$\mathbf{f}_{t+h} = \sum_{i=0}^{p-1} \hat{\Phi}_i \mathbf{y}_{t-i}, \tag{18}$$

Compared with univariate forecast models, the advantage of VAR models (same for all VAR-type of models which will follow) is that they provide predictions not only based on historical fitting of individual process but also by means of lags of other endogenous variables in the system. Note that, in our in-sample estimation and model selection, the maximum lag length considered will be $p = 12$.

5. **Bayesian Vector Autoregression** (BVAR)
While it is common to use VARs to obtain forecasts, it has also been argued that VARs estimated by Bayesian methods would provide better forecast with more parsimonious models because standard VARs often incur in over-fitting problems (Spencer, 1993). Compared with standard estimation, the BVAR[2] treats model's parameters as random variables, and applies Bayesian estimation imposing restrictions on the dynamics of the parameters according to a specific type of prior. Based on this assumption, the coefficients on longer lagged variables are more likely to be near zeros, resulting a more parsimonious estimation. Indeed, we still use the model showed in Equation (17), however, with the prior adopted being the Minnesota Prior (Del Negro and Schorfheide, 2004). In the VAR system, there are K equations, and each one can be expressed as,

$$y_{i,t+h-1} = \sum_{i=1}^{p} \sum_{j=1}^{K} \phi_{i,j} \cdot y_{j,t-i} + \epsilon_{i,t}, \tag{19}$$

In this case, the prior about coefficients are captured in the prior density function $g(\phi_{i,j})$. Then, using Bayesian theory, the estimators are obtained by the posterior density functions $g(\phi_{i,j}|y_{i,t})$.

$$g(\phi_{i,j}|y_{i,t}) = \frac{g(y_{i,t}|\phi_{i,j})g(\phi_{i,j})}{g(y_{i,t})}, \tag{20}$$

and the predictions of $y_{i,t}$ can be obtained from following,

$$f_{i,t+h} = \sum_{i=0}^{p-1} \sum_{j=1}^{K} \hat{\phi}_{i,j} \cdot y_{j,t-i}, \tag{21}$$

[2]For an application see Crompton and Wu (2005) who applied BVAR model to predict energy consumption in China.

where again, the maximum lag length considered will be $p = 12$.

6. **Factor Augmented Vector Autoregression** (FAVAR)
Last, we use a factor augmented VAR. The FAVAR model has been widely applied to large data especially in macroeconomics (Bernanke and Boivin (2003) and Bernanke et al. (2004)). Chudik and Pesaran (2011) claim that a less parameterised VAR model augmented with factors will not lose any relevant information and would often produce better forecasts than standard VARs. In the energy related literature, among others, Baumeister et al. (2016) have adopted VAR, BVAR and FAVAR models to predict gasoline price in US market.

The FAVAR aim at modelling a system $\mathbf{x}_t$ with N variables and assume a subset $\mathbf{y}$ of $\mathbf{x}_t$ which contains M variables, and the dynamics of $\mathbf{y}$ are driven by unobservable forces in $\mathbf{x}_t$. These unobservable forces are factors extracted from $\mathbf{x}_t$, containing most of the relevant information. The system can thus be formulated as follows:

$$\mathbf{x}_t = \mathbf{\Lambda}^f \cdot \mathbf{F} + \mathbf{\Lambda}^y \cdot \mathbf{y}_t + \epsilon_t, \tag{22}$$

where $\mathbf{\Lambda}^f$ is $N \times K$ coefficient matrix for K factors, $\mathbf{\Lambda}^y$ is a $N \times M$ coefficient matrix, and ϵ_t is a $N \times 1$ vector of error terms.

In this paper, we classify the energy demand processes into two groups, one group is the objective observed process for a specific source y_t, while the other group is made up of the energy processes obtained by other sources from which one factor is extracted. We denote this groups data as $\mathbf{x_t}$. The FAVAR model can be written in state-space form comprising two equations: the observation equation and the state equation. In the observation equation, the number K of factors F_t, where $K = 1$ in this paper, can be extracted from the variables in $\mathbf{x_t}$ through principal components. Thus, the state equation is,

$$\mathbf{z}_{\omega,t+h-1} = \sum_{i=1}^{p} \Phi_i \mathbf{z}_{\omega,t-i} + \epsilon_{\omega,t+h-1}, \tag{23}$$

where $\mathbf{z}_{\omega,t+h-1} = \begin{bmatrix} F_{\omega,t+h-1} \\ y_{\omega,t+h-1} \end{bmatrix}$. Again, Φ_i is a coefficient matrix, the dimension of which depends on the number of factors extracted from $\mathbf{x_t}$, and if there is only one factor, the coefficient matrix will be 2×2. Therefore, the objective process y_t can be predicted by one Φ_i that has been estimated,

$$\hat{\mathbf{z}}_{\omega,t+h} = \sum_{i=0}^{p-1} \hat{\Phi}_i \mathbf{z}_{\omega,t-i}, \tag{24}$$

where $\hat{\mathbf{z}}_{\omega,t+h} = \begin{bmatrix} \hat{F}_{\omega,t+h} \\ f_{\omega,t+h} \end{bmatrix}$. For each objective process, the predictions are obtained through the causal relationship with factors extracted from the remaining processes. In this paper, we extract one factor from remaining four energy series. For each energy source, the optimal model is selected by considering lag length k from 1 to 12.

3.2 IMS and DMS forecast methods

For all the models in each class, we construct forecasts of the demand for energy using both iterated multi-step (IMS) and direct multiple steps (DMS) methods. The IMS method provides h steps ahead predictions through a one step ahead predictor $f_1 = \hat{y}_{t+1}$ iterated forward h times. In each iteration, we estimate the Equation 25 below using the training sample, and then forecast one period ahead for the out-of-sample through Equation 26.

$$y_{t+1} = \boldsymbol{\Theta}_1(\mathcal{M}_i)y_t + \epsilon_{t+1}, \quad i = 1, \ldots, 6, \tag{25}$$

$$f_1 = \hat{y}_{t+1} = \hat{\boldsymbol{\Theta}}_1(\mathcal{M}_i)y_t, \tag{26}$$

where $\mathcal{M}_i$ is the ith model, and $\boldsymbol{\Theta}_1$ is a set of parameters in ith model.

The DMS method predicts h steps ahead through forecasting $f_h = \hat{y}_{t+h}$ directly. This is achieved by using the estimated Equation 27 using the training sample, and then using Equation 28 to predict f_h for the out-of-sample.

$$y_{t+h} = \boldsymbol{\Theta}_h(\mathcal{M}_i)y_t + \epsilon_{t+h}, \quad i = 1, \ldots, 6, \tag{27}$$

$$f_h = \hat{y}_{t+h} = \hat{\boldsymbol{\Theta}}_h(\mathcal{M}_i)y_t, \tag{28}$$

where $\boldsymbol{\Theta}_h$ is a set of parameters in ith model for DMS.

According to Marcellino et al. (2006), the IMS method should provide lower forecast error once the one-period ahead model is well specified. However, the DMS method is relatively more robust to misspecification in the forecast model. Thus, in general, DMS has been often preferred to IMS in empirical studies. As there is no strong evidence to support a clear cut choice between IMS and DMS, we shall obtain forecasts with both methods and will compare their respective forecasting ability.

3.3 Model Selection

For each model discussed above, we will consider different parameter settings and lag lengths, and assume that the best forecasting model exists among those considered here. The optimal or best model will be chosen in-sample by means of different information criteria including Akaike Information Criterion (AIC), Beyesian Information Criterion (BIC), Mallows' Information Criterion (MIC) (Hansen (2007) and Hansen (2008)) and the Jackknife (JKC), a cross-validation criterion suggested by Hansen and Racine (2012) and Hansen (2014). Each information criterion is computed using the in-sample fitted error $\hat{\epsilon}_t(m) = y_t - \hat{y}_t(m)$. Hence, the estimated fitted error variance equals to $\hat{\sigma}^2(m) = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}^2(m)$, where n is the number of observations in-sample.

The AIC and BIC information criteria reward for lower fitted errors but penalize for higher number of parameters estimated, so that $\hat{\sigma}^2(m)$ is the estimated error variance of model m, and the number of parameters in each estimated model is denoted as $k(m)$. The AIC and BIC can be respectively expressed as:

$$AIC = n \cdot ln(\hat{\sigma}^2(m)) + 2k(m), \tag{29}$$

$$BIC = n \cdot ln(\hat{\sigma}^2(m)) + k(m) \cdot ln(n), \tag{30}$$

where n is the total number of observations in-sample.

The Mallows' information criterion uses the estimated mean squared errors,

$$MIC = (y_t - \hat{y}_t(m))'(y_t - \hat{y}_t(m)) + 2 \cdot \hat{\sigma}^2(m) \cdot k(m), \tag{31}$$

where $\hat{y}_t(m)$ is the fitted value of y_t from model m and $\hat{\sigma}^2(m)$ and $k(m)$ are defined as above. Last, we use the Jackknife, a cross-validation criteria. To use this cross-validation method, we obtain a leave-one-out estimator for each in-sample point for every model, and then obtain a cross-validation fitted error $\tilde{\epsilon_t}$ through the following equation.

$$\tilde{e}_{m,i} = y_i - \hat{y}_{-i,m}, \tag{32}$$

where $\hat{y}_{-i,m}$ is the leave-one-out one step estimate of y_i based on the estimated parameters from the remaining observations. Then, the expression for the Jackknife (JKC) can be formulated as,

$$JKC = \frac{1}{n} \cdot \sum_{i=1}^{n} \tilde{\epsilon}_{m,i}^2, \tag{33}$$

Note that, for each of the criterion used, the optimal/best forecasting model will be the one which in each class/set of models minimizes the information criterion. In case of equal value of the information criterion for two different models within the same class, the more parsimonious model will be preferred.

3.4 Model Averaging

In this subsection, we briefly outline the model averaging methods which we shall use to improve the accuracy of our forecasts. We denote the prediction from i^{th} model as $f_t(i)$ for $1 \leq i \leq J$, and the average prediction f_t is defined as

$$f_t = \sum_{1}^{J} w_i f_t(i), \tag{34}$$

where $f_t(i)$ is obtained from the forecasting model class/set $\mathcal{M}_i$, and w_i is the weight attached to the individual $f_t(i)$ obtained from the J candidate forecasts. At this point, the major issue in model averaging becomes how to specify the weights w_i as different weighting functions are likely to provide different levels of forecasting accuracy.

3.4.1 Simple Model Averaging (SMA)

The simple model averaging provides an equally weighted average of the predictions of all the best forecast models in each class, such that the weights in simple model averaging are just $w_i = \frac{1}{J}$. The SMA forecast is,

$$f_t = \sum_{i=1}^{J} \frac{1}{J} f_t(i), \tag{35}$$

Note that SMA is known to improve the accuracy of forecasts as long as the model candidates are well specified. However, once some of the candidates are not well specified, the accuracy of averaged prediction will significantly decrease.

3.4.2 Granger-Ramanathan Model Averaging (GRMA)

Granger and Ramanathan (1984) proposed a model average weighting based on coefficients of the regression model. The regression is made of an average forecast f_t regressed on the candidate predictions $f_t(i)$ and is formulated as:

$$f_t = \beta_0 + \sum_{i}^{J} \beta_i f_t(i) + \epsilon_t$$

Granger and Ramanathan (1984) impose three constraints on the coefficients of this regression. First, the intercept coefficient is equal to zero, $\beta_0 = 0$; second, the coefficients on each candidate prediction should be non-negative, $\beta_i \geq 0$ for all i; last, the sum of coefficients of the regression must be equal to one, $\sum_1^J \beta_i = 1$. Having specified these constraints, they used the estimated coefficients as the weights for averaging, that is $w_i = \hat{\beta}_i$.

$$f_t = \sum_{i}^{J} \beta_i f_t(i), \tag{36}$$

3.4.3 Bayesian Model Averaging (BMA)

Bayesian model averaging assumes that there always exists at least one well-specified model among all candidate models and therefore one should give more weight to well or better specified candidates while less weight is attached to the rest of the models. The probability of candidate models to be well-specified is giving as a prior, and then the Bayesian posterior probability can be calculated conditional on real data. These Bayesian posterior probabilities for each candidate model are the weights for averaging all potential models. As the prior probability that any model is a well specified model is not known for each candidate model, the weights can be approximated by using the Bayesian information criteria.

$$w_m^b = \frac{exp(-\frac{1}{2}(BIC_m))}{\sum_{j=1}^{M} exp(-\frac{1}{2}BIC_j)}$$

Thus, the averaged forecast is given by:

$$f_t = \sum_1^J w_i^b f_t(i), \tag{37}$$

3.4.4 Other Model Averaging Functions

Anderson and Burnham (2002), proposed to replace BIC in the weighting function with AIC, this resulting into smoothed AIC (SAIC) weighting function. In this case the weights w_m^a will be:

$$w_m^a = \frac{exp(-\frac{1}{2}(AIC_m))}{\sum_{j=1}^{M} exp(-\frac{1}{2}AIC_j)}$$

and the SAIC model averaging (AMA) forecasts are given by:

$$f_t = \sum_1^J w_i^a f_t(i), \tag{38}$$

Similarly, as suggested by Hansen (2007), Hansen (2008) and (Hansen and Racine (2012), Hansen (2014) we can use Mallows' information criteria (MIC) or the Jackknife cross-validation criteria (JKC) to replace the BIC, thus obtaining weights w_m^m and w_m^J. such that the two weighting functions will respectively be,

$$w_m^m = \frac{exp(-\frac{1}{2}(MIC_m))}{\sum_{j=1}^{M} exp(-\frac{1}{2}MIC_j)}$$

with the Mallows' model averaging (MMA) forecast being,

$$f_t = \sum_1^J w_i^m f_t(i), \tag{39}$$

and

$$w_m^j = \frac{exp(-\frac{1}{2}(JKC_m))}{\sum_{j=1}^{M} exp(-\frac{1}{2}JKC_j)}$$

while the Jackknife model averaging (JMA) forecast is,

$$f_t = \sum_1^J w_i^j f_t(i), \tag{40}$$

4 Results and Comparison of the Forecasts

In this section, we provide our results and a comparison between forecast of different time horizons according to the various criteria. Because of the heavy computational burden, we shall consider static forecasts obtained with both IMS and DMS methods. Our forecast methods can be extended to a dynamic type by constructing recursive or rolling out-of-sample forecasts iteratively, however, this would be made at the expense of an even heavier computation burden and it is not done here.

In the following, we report results of, and comparisons between different forecast models, forecast methods, forecast model selection criteria and eventually model averaging methods. The accuracy of these predictors is measured by Mean Squared Forecast Error (MSFE hereafter), and model averaging predictions are statistically tested by using tests provided by Diebold and Mariano (1995). We predict demand for each energy source with forecast horizons of 30 minutes, 1 hour, 2 hours, 4 hours, 8 hours, 12 hours, 18 hours and 24 hours.

4.1 Comparison based on Forecast Error

For any forecast model i, if $f_t(i)$ is the predicted value of objective y_t, and the forecast error $\hat{\epsilon}_t(i)$ is expressed as,

$$\hat{\epsilon}_t(i) = y_t - f_t(i), \tag{41}$$

Thus, the estimated forecast error variance is,

$$\hat{\sigma}^2(i) = \frac{1}{m}\sum_{i=1}^{m} \hat{\epsilon}_i(i)^2, \tag{42}$$

where m is the number of out-of-sample predicted points. This paper uses the MSFE to measure the accuracy of predictions.

$$MSFE(m) = E[y_t - f_t(m)]^2, \tag{43}$$

Table 2 displays the MSFE of an average between IMS and DMS forecasts of the best model of each model class. There are several points to note. In terms of the model selection criteria, we notice that the best/optimal models suggested by the JKC tends to often produce more precise forecasts in terms of lower forecast error. Comparing different classes of models, NARNN and VAR appear to forecast better than the others model classes. Within the five energy sources, we can see that it is easier to obtain accurate predictions of the demand for nuclear energy. This result is fairly expected as nuclear plants once on-line run flat-out and also confirms the preliminary analysis above which showed less noise in nuclear energy demand process. On the contrary, CCGT and wind energy demands are relatively more difficult to predict, and display larger values of the MSFE.

Table 3 reports the MSFE of the different model averaging forecasts. Although the DMS method reports lower MSFE in 16 out of the 30 cases, it is still hard to conclude whether IMS beats DMS or the reverse, because each method dominates the other depending on the source of energy for which demand is forecast. Among the six types of model averaging methods, generally the BMA, MMA are superior to the others. Lastly but most importantly, from Table 3, we can see that there always exists a model averaging method which gives a lower MSFE than the best/optimal model from any class as from Table 2.

Insert Table [2] and Table [3] about Here

4.2 IMS vs DMS

In this subsection, we specifically compare the forecasting ability of IMS and DMS. Figure 6 shows the plots of the MSFEs obtained from the two forecasting methods. Generally, it seems preferable to use DMS, particularly if using ARMA and FAVAR models to forecast a longer horizon. An exception is the case of BVAR for which the IMS beats the DMS in generating forecasts of the demand for energy produced by coal, ccgt, wind and hydro-power.

In more detail, the ARMA model with DMS method produces better forecasts for coal, nuclear and hydro-power. With regard to CCGT and wind sources, the FAVAR shows better forecasting ability, but with IMS in one instance and with DMS in another. Also, as a general result, we observe that the predictions are more accurate in the beginning and at the end of the forecast horizon, as shown by the MSFE, which is always low in forecasting 30 minutes, 1 hour and 1 day ahead, but grows to higher levels in the mid-term, indicating that serial correlation or cross-serial correlation (in multivariate models) is stronger in the short and long-term but weaker in the medium term (relative to the frequency of the data). Therefore, for empirical purposes, forecasting with DMS is advised for short and longer terms.

Insert Figure [6] about Here

4.3 The Comparison of Information Criteria

In section 3.3 we outlined four types of information criteria used to selecting the optimal in-sample forecast model. Here we compare their performance out-of-sample, computing the MSFE of the up-to-one-day forecasts of the demand for energy produced from the best models of each set as selected by each information criteria. Figure 7 illustrates he MSFE of the optimal forecast models in each class as selected by AIC, BIC, MIC and JKC respectively.

In brief, BIC and MIC consistently suggest the same optimal forecast model, while the AIC and JKC are more likely to indicate similar forecast model. For all energy sources, the MSFE obtained from the forecasting model suggested by AIC and JKC almost never underperform to the counterpart suggested by BIC and MIC. Particularly in each sector, the optimal models suggested by these four information criteria are more or less the same for HW, NARNN, VAR and BVAR models, and the only exception is the BVAR model for wind-produced energy. In term of ARMA and FAVAR models, the AIC and JKC select better out-of-sample forecasting model, except for the short horizon forecasts generated by ARMA model in the coal-produced energy demand series, and relatively longer horizons by ARMA for hydro-power. Now, since each combination method is averaging the optimal models from six types of forecasting candidates, the fact that AIC and JKC are selecting models that give lower MSFE, it is reasonable to expect that using these criteria in weighting function for model averaging will also give more accurate predictions. Below we examine the issue in more detail.

Insert Figure [7] about Here

4.4 Comparison of Model Averaging Methods

In this sub-section, we compare the forecasting performance of the different model averaging methods, namely: SMA, GRMA, AMA, BMA, MMA and JMA. Figure 8 displays the MSFEs obtained out-of-sample for the various model averaging methods. The multi-steps predictions are computed with both IMS and DMS. Consistent with the results displayed in Table 3 and Figure 6 and already discussed, the DMS method provides slightly more accurate forecasts than the IMS.

Within the IMS-based forecasts, we find that, overall, SMA and AMA produce more accurate predictions for nuclear, CCGT and wind for most of the forecast horizons, while MMA is superior in generating predictions for coal and hydro. In more detail, the AMA, BMA and MMA methods produce good forecasts albeit none of them clearly dominated the others. Also, the JMA generates better forecasts for coal but loses efficiency as the forecast horizon approaches the 24 hours; the BMA method produces better predictive ability for nuclear and CCGT, while MMA is best for hydro-power sourced energy; the AMA provides better forecasts for the demand of wind-sourced energy. On another hand, if one used DMS, the AMA, BMA, MMA and JMA show very similar forecast abilities with slightly differences for the demand produced by nuclear, CCGT and hydro-power sources. Although the performance of the model averaging methods in forecasting the demand for energy obtained using coal and hydro-power are not much different regardless of the weighting functions used, the MMA and AMA produce slightly better predictions for the demand of energy sourced from coal and wind, respectively.

In more depth, in the case of coal-based energy, the GRMA predictors obtain lower forecast error in the beginning of the predicted period and at its end, while JMA often predictions are superior in the middle of the forecast period. For nuclear-sourced energy, the BMA consistently outperform to other model averaging methods in the accuracy of its prediction. For CCGT, all model averaging techniques perform more or less the same, and again, BMA is the one that slightly more accurate than the others. Regarding the demand for wind-sourced energy, the MMA provides the most accurate predictions in forecasting the longer term (one-day ahead), but AMA generally outperforms the others. Lastly, considering hydro-power produced energy, using IMS, the MMA shows the best forecasting ability while GRMA performs poorly. However, using DMS, both MMA and GRMA produce more accurate forecasts compared with rest of the model averaging methods.

Insert Figure [8] about Here

Next, we use the Diebold-Mariano (DM) and the Wilcoxon's sign-ranked (Sign) tests by (Diebold and Mariano, 1995) to test the prediction equivalence of AMA, BMA, MMA and JMA (given the relatively poorer performance of SMA and GRMA). Table 4 displays the pair-wise results of both tests for IMS and DMS, respectively. Generally, both the DM and the Sign test provide similar results, expect for a few cases which concern the predictions obtained by IMS. On Table4, for a pair of ordered forecasts obtained by weight functions x and y, a positive (negative) and statistically significant value of the statistic would imply superiority (inferiority) of the predictions obtained with weighting function y over those obtained if using weighting function x. The results can be roughly summarized as: AMA and BMA provide statistically equivalent forecasts for all energy sectors regardless of whether the forecasts are obtained by means of IMS or DMS methods. When using IMS, the MMA forecasts are, in general, found statistically superior to the AMA/BMA except for Nuclear Energy where contrasting results are found. Also, JMA-produced forecasts, although often equivalent to AMA and BMA, are found in a few cases to be even slightly superior to them. For DMS forecasts, the MMA again outperforms all others methods while the JMA loses its slight superiority compared to the AMA and BMA forecasts and it is actually outperformed when producing forecasts for coal, nuclear and hydro generated energy. Finally, for both the IMS and DMS methods, MMA obtained forecasts are found consistently superior to those given by the JMA weight function.

Insert Table [4] about Here

5 Forecasting the Demand for Energy in Levels

In this section, we finally obtain forecast of the levels of the energy demand for the five energy sources on 22 March 2016. Specifically, we re-combine the IMS and DMS predictions obtained through all the model averaging methods, with the deterministic/periodic terms captured by Equations 4, 8 and 9, therefore obtaining forecasts for the level of the UK demand for energy produced by the five sources considered. Figure 9 illustrate the predicted and actual values for coal, nuclear, CCGT, wind and hydro-power, respectively. Treating the forecasts from the individual forecast models as bench-marks, the figures compare the predictions obtained from model averaging to that of the individual forecasting models obtained by means of both IMS and DMS.

Among six bench-mark forecast models, the HWS model is the worst performing one, and the NARNN is the best and actually shows a performances close to that of model averaging methods. However, it is important to re-iterate that there always exists a model averaging forecast that can beat the forecasts obtained from a bench-mark model. Another notable fact is that the prediction become more accurate as the forecast horizon approaches the 24 hours, and also that DMS outperform IMS in longer-term forecasting.

In more detail, the model averaging predictions for coal, CCGT, wind and hydro-power, remain relatively accurate after adding the periodic and deterministic components. This is not the case for the demand for energy nuclear-sourced, where in fact, forecasts of the level is less accurate due to the inaccuracy in fitting the deterministics.

Using the IMS method, predictions become more and more accurate reaching the 1 day horizon. For the coal-sourced energy, to predict the shorter-term (30 minutes - 4 hours), the GRMA method produces the most precise forecasts, while MMA becomes superior in forecasting the longer horizon (6 - 24 hours). For nuclear energy, MMA and JMA allow us to obtain better predictions than others. All model averaging methods give similar forecasting of the CCGT-sourced energy. AMA predictions outperform the other forecasting methods for the the demand of wind-fuelled energy, and lastly, in the hydro-power sector, the MMA method again shows best forecasting ability than other model averaging methods.

Insert Figure [9] about Here

6 Conclusions

In this paper we have produced more accurate short-term forecasts of the demand for energy in the UK using a forecasting approach based on model averaging of several popular linear or non-linear, univariate and multivariate forecast models. Specifically, we used an algorithm that once obtained the forecasts from sets of ARMA, Holt-Winters, Non Linear Autoregressive Neural Networks, Vector Autoregressions, Bayesian VAR and Factor Augmented VAR models selects the best forecasting model from each model-set according to four different information criteria (AIC, BIC, Mallows' and Jackknife). The best models as selected by each of the different information criteria within each model set are then averaged using six different combination weight metrics including Simple Model Averaging, Granger-Ramanathan Model Averaging, Bayesian Model Averaging, Akaike Model Averaging, Mallows Weights and the Jackknife.

Our results confirm the merits of combination forecasting as a superior forecasting strategy. Among the single forecasting models, NARNN and VAR forecast are superior in terms of lower MSFE whilst HWS perform worst. Unexpectedly, DMS forecasts outperforms those obtained by IMS in terms of accuracy. For all energy sources, the MSFE obtained from the forecasting model selected by AIC and JKC almost never underperform compared to their counterparts suggested by BIC and MIC. Among the six types of model averaging methods, generally the BMA, MMA are superior to the others. Lastly but most importantly, there always exists a model averaging method which gives a lower MSFE than the best/optimal models within each class however selected.

As highlighted above, accurate forecasts are a precious resource for demand response, and/or load management. With timely and accurate prediction of demand, load management programs facilitate system load balancing by avoiding peak occurrences. On the other hand, they can also be crucial for demand response, which has been gaining prominence in recent years as an effective and inexpensive tool for reducing overall utility peak demand while improving system-wide energy efficiency. Through the curtailment of electricity consumed by end-users during periods of high demand or electricity grid instability, demand response technology addresses unexpected variances in electricity supply and demand levels. When wholesale electricity market prices are high or when overall grid system reliability is compromised, demand response programs offer incentives to end-users in order to affect time of use, instantaneous demand level, and/or aggregate electricity consumption.

Given accurate forecasts with low error such as those we obtained using model averaging, it is theoretically possible for network management to, for example, temporarily curtail a portion of the network load in some areas of a city whenever approaching a predetermined peak in demand in others areas. It is thus necessary for the network management to determine an acceptable peak demand load maximum for the various areas. The real-time energy monitoring system provided by smart metering together with the model averaging forecast signals an upcoming breach of the predetermined peak demand load maximum. Thus, curtailment policy shreds unnecessary loads during these events in order to control overall peak loading and prevent an unwanted peak demand occurrence. Clearly, accurate model averaging forecasts would be particularly useful for efficient and cost-effective peak demand energy management across city municipalities and other large energy end-users. In this case there would be added benefit not only to the electric utility provider but also to the environment through efficient and reduced power generation capacity. Such reduction and efficient usage of power generation would undoubtedly contribute to the energy sustainability of local municipalities and their communities.

References

Amjady, N. (2001), 'Short-term hourly load forecasting using time-series modeling with peak load estimation capability', *IEEE Transactions on Power Systems* **16**(3), 498–505.

Anderson, D. R. and K. P. Burnham (2002), 'Avoiding pitfalls when using information-theoretic methods', *The Journal of Wildlife Management* pp. 912–918.

Badri, M. A., A. Al-Mutawa, D. Davis and D. Davis (1997), 'Edssf: a decision support system (dss) for electricity peak-load forecasting', *Energy* **22**(6), 579–589.

Bates, John M and Clive WJ Granger (1969), 'The combination of forecasts', *Journal of the Operational Research Society* **20**(4), 451–468.

Baumeister, C., L. Kilian and T. K. Lee (2016), 'Inside the crystal ball: New approaches to predicting the gasoline price at the pump', *Journal of Applied Econometrics* .

Bernanke, B. S. and J. Boivin (2003), 'Monetary policy in a data-rich environment', *Journal of Monetary Economics* **50**(3), 525–546.

Bernanke, B. S., J. Boivin and P. Eliasz (2004), Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach, Technical report, National Bureau of Economic Research.

Chow, TWS and CT Leung (1996), 'Neural network based short-term load forecasting using weather compensation', *IEEE Transactions on Power Systems* **11**(4), 1736–1742.

Chudik, A. and M. Pesaran (2011), 'Infinite-dimensional vars and factor models', *Journal of Econometrics* **163**(1), 4–22.

Crompton, P. and Y. Wu (2005), 'Energy consumption in china: past trends and future directions', *Energy economics* **27**(1), 195–208.

Cuestas, J. C. and L. A. Gil-Alana (2016), 'Testing for long memory in the presence of non-linear deterministic trends with chebyshev polynomials', *Studies in Nonlinear Dynamics & Econometrics* **20**(1), 57–74.

De Gooijer, Jan G and Kuldeep Kumar (1992), 'Some recent developments in non-linear time series modelling, testing, and forecasting', *International Journal of Forecasting* **8**(2), 135–156.

Del Negro, Marco and Frank Schorfheide (2004), 'Priors from general equilibrium models for vars', *International Economic Review* **45**(2), 643–673.

Diebold, Francis X and Roberto S Mariano (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.

Ediger, Volkan Ş and Sertac Akar (2007), 'Arima forecasting of primary energy demand by fuel in turkey', *Energy Policy* **35**(3), 1701–1708.

Ediger, Volkan Ş, Sertaç Akar and Berkin Uğurlu (2006), 'Forecasting production of fossil fuel sources in turkey using a comparative regression and arima model', *Energy Policy* **34**(18), 3836–3846.

Ermis, K, A Midilli, I Dincer and MA Rosen (2007), 'Artificial neural network analysis of world green energy use', *Energy Policy* **35**(3), 1731–1743.

Fan, JY and JD McDonald (1994), 'A real-time implementation of short-term load forecasting for distribution power systems', *IEEE Transactions on Power Systems* **9**(2), 988–994.

Filik, Ümmühan Başaran, Ömer Nezih Gerek and Mehmet Kurban (2011), 'A novel modeling approach for hourly forecasting of long-term electric energy demand', *Energy Conversion and Management* **52**(1), 199–211.

Fouquet, Roger, Peter Pearson, David Hawdon, Colin Robinson and Paul Stevens (1997), 'The future of uk final user energy demand', *Energy Policy* **25**(2), 231–240.

Francis, Brian M, Leo Moseley and Sunday Osaretin Iyare (2007), 'Energy consumption and projected growth in selected caribbean countries', *Energy Economics* **29**(6), 1224–1232.

García-Ascanio, Carolina and Carlos Maté (2010), 'Electric power demand forecasting using interval time series: A comparison between var and imlp', *Energy Policy* **38**(2), 715–725.

Geem, Zong Woo and William E Roper (2009), 'Energy demand estimation of south korea using artificial neural network', *Energy policy* **37**(10), 4049–4054.

Granger, Clive WJ and Ramu Ramanathan (1984), 'Improved methods of combining forecasts', *Journal of Forecasting* **3**(2), 197–204.

Guidolin, Massimo and Allan Timmermann (2007), 'Asset allocation under multivariate regime switching', *Journal of Economic Dynamics and Control* **31**(11), 3503–3544.

Haas, Reinhard and Lee Schipper (1998), 'Residential energy demand in oecd-countries and the role of irreversible efficiency improvements', *Energy economics* **20**(4), 421–442.

Hagan, Martin T and Suzanne M Behr (1987), 'The time series approach to short term load forecasting', *IEEE Transactions on Power Systems* **2**(3), 785–791.

Hansen, Bruce E (2007), 'Least squares model averaging', *Econometrica* **75**(4), 1175–1189.

Hansen, Bruce E (2008), 'Least-squares forecast averaging', *Journal of Econometrics* **146**(2), 342–350.

Hansen, Bruce E (2014), 'Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation', *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics, forthcoming* .

Hansen, Bruce E and Jeffrey S Racine (2012), 'Jackknife model averaging', *Journal of Econometrics* **167**(1), 38–46.

Hendry, David F and Michael P Clements (2004), 'Pooling of forecasts', *The Econometrics Journal* **7**(1), 1–31.

Hong, Wei-Chiang (2013), *Intelligent energy demand forecasting*, Springer.

Hunt, Lester C, Guy Judge and Yasushi Ninomiya (2003), 'Underlying trends and seasonality in uk energy demand: a sectoral analysis', *Energy Economics* **25**(1), 93–118.

Kurban, Mehmet and U Basaran Filik (2009), 'Next day load forecasting using artificial neural network models with autoregression and weighted frequency bin blocks', *International Journal of Innovative Computing, Information and Control* **5**(4), 889–898.

Lai, TM, WM To, WC Lo and YS Choy (2008), 'Modeling of electricity consumption in the asian gaming and tourism centerâĂŤmacao sar, people's republic of china', *Energy* **33**(5), 679–688.

Madigan, David and Adrian E Raftery (1994), 'Model selection and accounting for model uncertainty in graphical models using occam's window', *Journal of the American Statistical Association* **89**(428), 1535–1546.

Maia, Andre Luiz S, Francisco de AT de Carvalho and Teresa B Ludermir (2006), Symbolic interval time series forecasting using a hybrid model, *in* '2006 Ninth Brazilian Symposium on Neural Networks (SBRN'06)', IEEE, pp. 202–207.

Marcellino, Massimiliano, James H Stock and Mark W Watson (2006), 'A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series', *Journal of econometrics* **135**(1), 499–526.

Markham, Ina S and Terry R Rakes (1998), 'The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression', *Computers & operations research* **25**(4), 251–263.

McAvinchey, Ian D and Andreas Yannopoulos (2003), 'Stationarity, structural change and specification in a demand system: the case of energy', *Energy Economics* **25**(1), 65–92.

Nogales, Francisco Javier, Javier Contreras, Antonio J Conejo and Rosario Espínola (2002), 'Forecasting next-day electricity prices by time series models', *IEEE Transactions on power systems* **17**(2), 342–348.

Pao, Hsiao-Tien (2006), 'Comparing linear and nonlinear forecasts for taiwan's electricity consumption', *Energy* **31**(12), 2129–2141.

Sadorsky, Perry (2009), 'Renewable energy consumption, co 2 emissions and oil prices in the g7 countries', *Energy Economics* **31**(3), 456–462.

Sözen, Adnan, Erol Arcaklioğlu and Mehmet Özkaymak (2005), 'TurkeyâĂŹs net energy consumption', *Applied Energy* **81**(2), 209–221.

Spencer, David E (1993), 'Developing a bayesian vector autoregression forecasting model', *International Journal of Forecasting* **9**(3), 407–421.

Sumer, Kutluk Kagan, Ozlem Goktas and Aycan Hepsag (2009), 'The application of seasonal latent variable in forecasting electricity demand as an alternative method', *Energy policy* **37**(4), 1317–1322.

Timmermann, Allan (2006), 'Forecast combinations', *Handbook of economic forecasting* **1**, 135–196.

Weron, Rafal (2007), *Modeling and forecasting electricity loads and prices: A statistical approach*, Vol. 403, John Wiley & Sons.

Zhang, G Peter (2003), 'Time series forecasting using a hybrid arima and neural network model', *Neurocomputing* **50**, 159–175.

A Appendix: Prediction Equivalent Tests

Diebold and Mariano (1995) proposed statistical tests to compare the forecasting errors from pairwise models. In the present paper, we introduce two types of tests: Diebold and Mariano asymptotic test and Wilcoxon's signed-rank test. These two tests are aiming to distinguish the null that

$$H_0: \quad E[g(e_{i,t})] = E[g(e_{j,t})]$$

versus,

$$H_1: \quad E[g(e_{i,t})] \neq E[g(e_{j,t})]$$

where $g(e_{i,t})$ is a forecasting loss function on model i. Also, define that the loss differential series $d_t \equiv [g(e_{i,t}) - g(e_{j,t})]$ for model i and j. Thus, hypothesis can also be understood as $E[d_t] = 0$.

The first test used is Diebold and Mariano asymptotic test, which is under mild assumption that d_t is a covariance stationary and short memory series. Then, we have,

$$S_1 = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{f}_d(0)}{T}}} \overset{a}{\sim} N(0,1) \tag{44}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}[g(e_{i,t}) - g(e_{j,t})]$, and the variance term,

$$2\pi \hat{f}_d(0) = \sum_{\tau=-(T-1)}^{(T-1)} I(\frac{\tau}{S(T)})\hat{\gamma}_d(\tau)$$

where $I(\frac{\tau}{S(T)})$ is the lag window and $S(T)$ is the truncation lag. Noted that $I(\frac{\tau}{S(T)}) = 0$ for $|\tau| > h-1$ as the h-step-ahead forecast errors are $h-1$ dependent at most.

$$\hat{\gamma}_d(\tau) = \frac{1}{T}\sum_{t=|\tau|+1}^{r}(d_t - \bar{d})(d_{t-|\tau|} - \bar{d})$$

As the Diebold and Mariano Test is a two-side test, it not only tests the equivalence, but also provides superior and inferior comparisons. A case that the statistic S_1 falls outside of the right(left)-hand confidence interval implies the forecasting error $e_{i,t}$ ($e_{j,t}$) is greater than $e_{j,t}$ ($e_{i,t}$) with a measurable function $g(\cdot)$, thus, the predictor $f_{i,t}$ ($f_{j,t}$) is less accurate than $f_{j,t}$ ($f_{i,t}$).

The second test introduced is Wilcoxon's signed-rank test. The test statistics follows a standard normal distribution under the assumption that loss differential series d_t is independent identically distributed (i.i.d). Since we compare predictors for different forecasting horizons, the d_t is reasonable to be i.i.d.

$$S_2 = \frac{S_{1a} - \frac{T(T+1)}{4}}{\sqrt{\frac{T(T+1)(2T+1)}{24}}} \overset{a}{\sim} N(0,1) \tag{45}$$

where

$$S_{2a} = \sum_{t=1}^{T} I_+(d_t) rank(|d_t|)$$

where $I_+(d_t) = 1$ if $d_t > 0$ and it equals to 0 otherwise. The $rank(\cdot)$ is the Wilcoxon's rank operator. Wilcoxon's signed-rank test can also compare the superiority-inferiority through the sign.

B Tables

Table 2: MSFE of Individual Models

This table reports the mean square forecast error of the best individual forecast models according to different information criteria. The MSFE averages those of IMS and DMS predictions from all forecast horizons.

		ARMA	HW	NARNN	VAR	BVAR	FAVAR
Coal	AIC	0.0428	1.3233	0.0284	0.0304	0.0438	0.0274
	BIC	0.0428	1.3233	0.0284	0.0305	0.0436	0.0334
	MIC	0.0428	1.3706	0.0284	0.0327	0.0429	0.0397
	JKC	0.0289	1.3233	0.0295	0.0304	0.0438	0.0274
Nuclear	AIC	0.0239	0.0005	0.0005	0.0004	0.0013	0.0007
	BIC	0.0239	0.0005	0.0004	0.0004	0.0012	0.0016
	MIC	0.0239	0.0005	0.0005	0.0004	0.0012	0.0007
	JKC	0.0004	0.0005	0.0005	0.0004	0.0013	0.0007
CCGT	AIC	0.1278	0.1527	0.0971	0.0935	0.1378	0.1017
	BIC	0.1278	0.1527	0.0963	0.0954	0.1404	0.1013
	MIC	0.1278	0.1565	0.1032	0.0955	0.1480	0.0962
	JKC	0.1083	0.1527	0.0971	0.0935	0.1378	0.1017
Wind	AIC	0.1564	0.1201	0.0442	0.0909	0.1254	0.0835
	BIC	0.1742	0.1201	0.0545	0.0866	0.1384	0.4522
	MIC	0.1564	0.0843	0.0442	0.0860	0.1898	0.0835
	JKC	0.1097	0.1201	0.0442	0.0909	0.1254	0.0835
Hydro-Power	AIC	0.0508	0.7546	0.0591	0.0574	0.0483	0.0576
	BIC	0.0508	0.7546	0.0571	0.0553	0.0490	0.0586
	MIC	0.0487	0.6072	0.0591	0.0547	0.0523	0.0588
	JKC	0.1790	0.7546	0.0591	0.0574	0.0573	0.0576

Table 3: MSFE of Model Averaging Methods

This table documents the mean square forecast error of the model average methods. The MSFE averages predictors from all forecast horizons.

		SMA	GRMA	AMA	BMA	MMA	JMA
Coal	IMS	0.0440	0.0440	0.0297	0.0297	0.0231	0.0340
	DMS	0.0496	0.0330	0.0268	0.0268	0.0255	0.0430
Nuclear	IMS	0.0013	0.00022	0.0005	0.0004	0.00022	0.0007
	DMS	0.00038	0.00063	0.00043	0.00047	0.00036	0.00036
CCGT	IMS	0.1004	0.1275	0.0932	0.0927	0.0988	0.0996
	DMS	0.1113	0.1191	0.1017	0.1004	0.1175	0.1086
Wind	IMS	0.0805	0.2896	0.0429	0.0457	0.0821	0.0679
	DMS	0.0508	0.1366	0.0390	0.0423	0.0536	0.0584
Hydro-Power	IMS	0.0606	0.0958	0.0500	0.0474	0.0449	0.0543
	DMS	0.0635	0.0426	0.0508	0.0508	0.0483	0.0556

Table 4: Prediction Equivalent Tests on Model Average Predictors

This table reports the prediction equivalence results for pair-wise model averaging methods which are using the IMS and DMS. DM_test refers to the Diebold and Mariano test, and Sign_test refers to the Wilcoxon's signed-rank test. Both tests use the critical values of standard normal distributions.

	IMS									
	Coal		Nuclear		CCGT		Wind		Hydro	
	DM_test	Sign_test	DM_test	Sign_test	DM_test	Sign_test	DM_test	Sign_test	DM_test	Sign_test
AMA vs BMA	-0.78(0.42)	-0.42(0.67)	-0.83(0.34)	-1.40(0.18)	1.42(0.16)	2.10(0.04)	0.76(0.42)	1.52(0.14)	-0.93(0.35)	1.12(0.26)
AMA vs MMA	3.62(0.00)	2.52(0.01)	-0.48(0.63)	1.68(0.09)	1.35(0.18)	2.10(0.04)	1.56(0.12)	2.10(0.04)	1.93(0.05)	2.10(0.04)
AMA vs JMA	-1.65(0.10)	1.12(0.26)	1.74(0.08)	2.52(0.01)	-0.33(0.74)	1.68(0.09)	0.50(0.62)	2.10(0.04)	-0.97(0.33)	1.12(0.26)
BMA vs MMA	3.63(0.00)	2.52(0.01)	-0.27(0.78)	2.38(0.02)	-0.56(0.58)	1.68(0.09)	-1.21(0.23)	1.12(0.26)	5.36(0.00)	2.52(0.01)
BMA vs JMA	-1.65(0.10)	1.12(0.26)	2.76(0.01)	2.52(0.01)	-1.00(0.32)	1.12(0.26)	-1.89(0.06)	0.42(0.67)	-0.91(0.36)	1.12(0.26)
MMA vs JMA	-2.77(0.01)	-2.52(0.01)	0.71(0.48)	1.12(0.26)	-1.91(0.06)	-0.42(0.68)	-1.35(0.18)	0.42(0.67)	-1.20(0.23)	1.12(0.26)
	DMS									
	Coal		Nuclear		CCGT		Wind		Hydro	
AMA vs BMA	-0.38(0.70)	2.10(0.04)	0.59(0.55)	0.42(0.67)	-0.59(0.55)	0.42(0.67)	1.57(0.12)	2.10(0.04)	-0.53(0.59)	1.12(0.26)
AMA vs MMA	6.12(0.00)	2.52(0.01)	-2.22(0.03)	-0.42(0.67)	2.54(0.01)	2.10(0.04)	2.31(0.02)	2.38(0.02)	0.91(0.36)	2.10(0.04)
AMA vs JMA	-2.01(0.04)	-0.42(0.67)	-2.15(0.03)	-1.40(0.18)	-0.74(0.46)	1.12(0.26)	2.94(0.00)	2.52(0.01)	-2.33(0.02)	-0.42(0.67)
BMA vs MMA	6.13(0.00)	2.52(0.01)	-2.20(0.03)	0.42(0.67)	2.57(0.01)	2.10(0.04)	1.99(0.05)	2.38(0.02)	1.84(0.07)	2.38(0.02)
BMA vs JMA	-2.01(0.04)	-0.42(0.67)	-2.07(0.04)	-1.40(0.18)	-0.57(0.57)	1.68(0.09)	2.83(0.00)	2.38(0.02)	-2.35(0.02)	-0.42(0.67)
MMA vs JMA	-4.03(0.00)	-2.52(0.01)	-0.91(0.36)	1.12(0.26)	-3.46(0.00)	-1.40(0.18)	-0.88(0.38)	1.12(0.26)	-2.79(0.01)	-0.42(0.67)

C Figures

Figure 2: **The levels of energy demand in UK**

This figure plots the level of demand for each energy source.

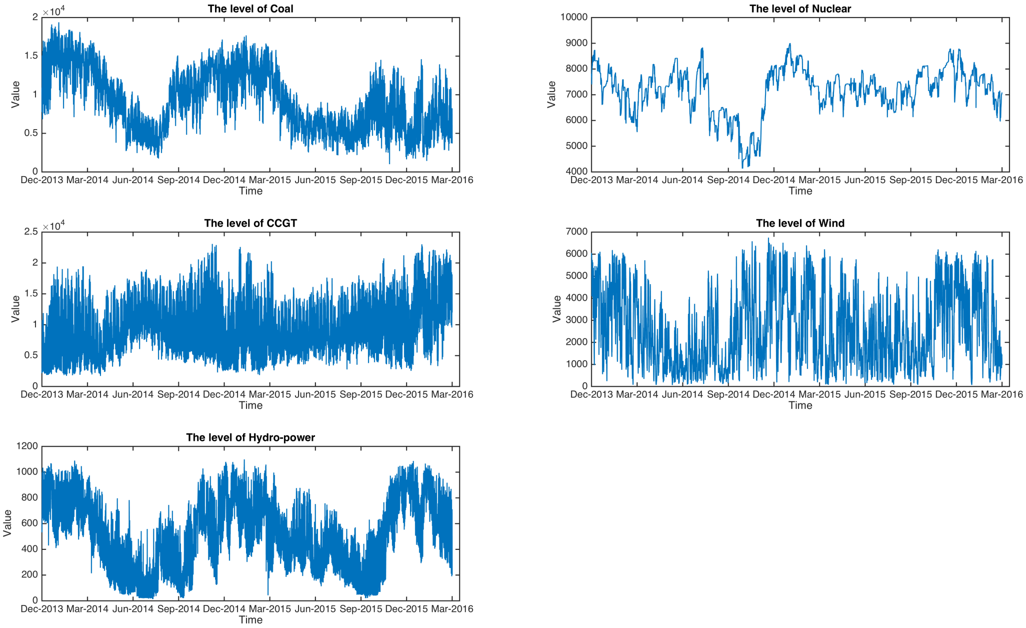


Figure 3: **The seasonal component in the demand for energy**

This figure plots deterministic seasonal trend fitted by Chebyshev polynomials with order 5 for each energy source.

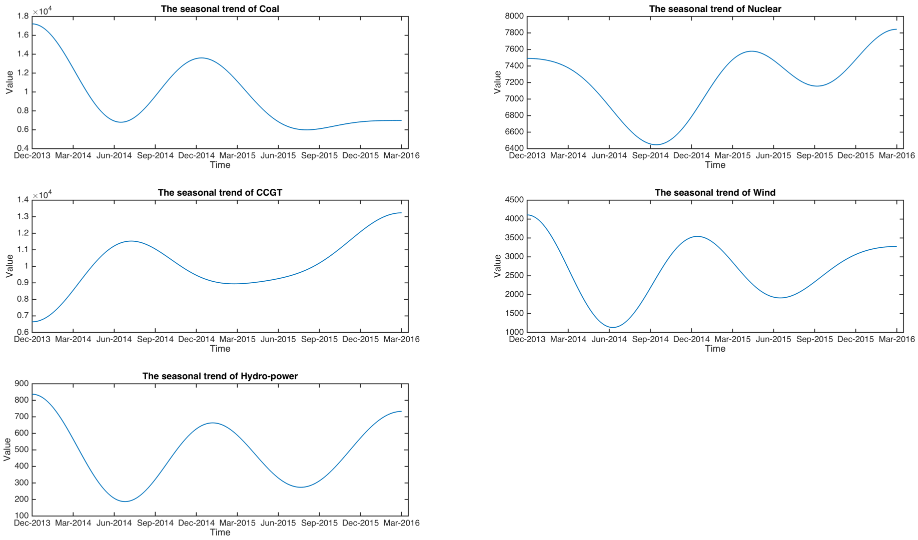


Figure 4: **The day of the week component in the demand for energy**

This figure plots deterministic weekly trend fitted by Weron (2007)'s moving average method with cyclical length equals to 336. Thus, the weekly trend contains 336 observations, which indicates the cyclical patterns for one week.

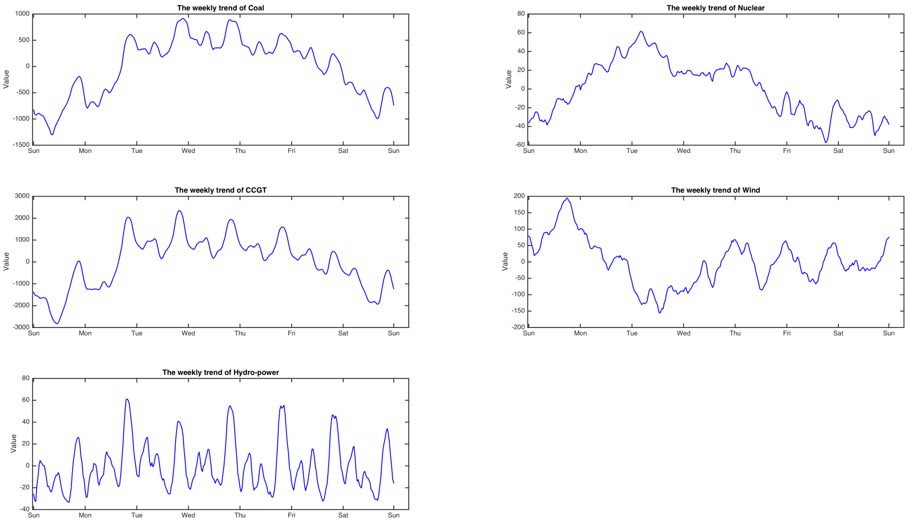


Figure 5: **The Stochastic Component of Demand for Energy**

This figure plots stochastic term after remove deterministic non-linear patterns for each energy source.

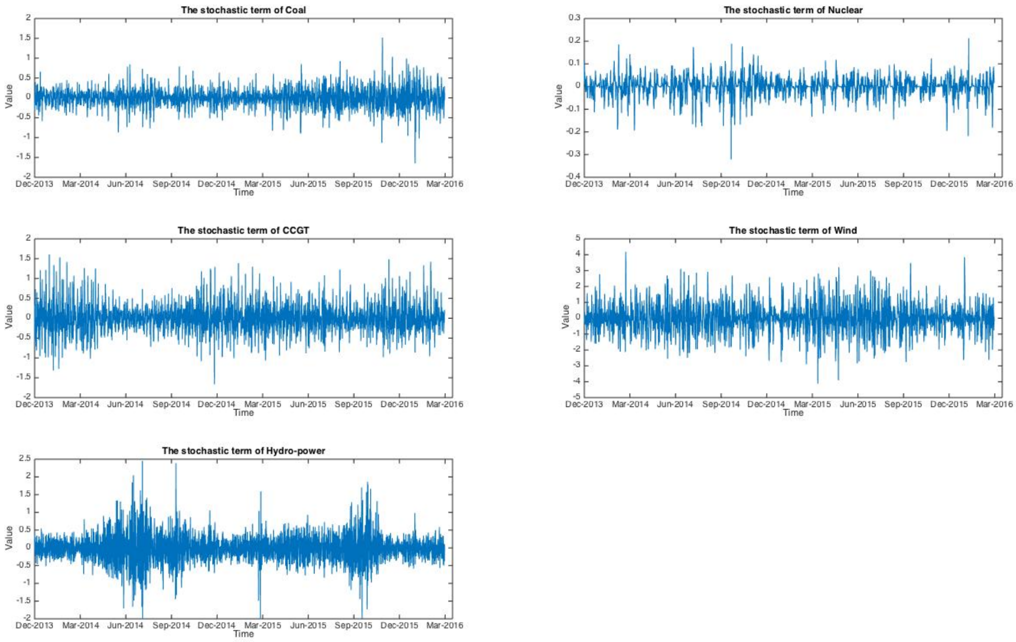
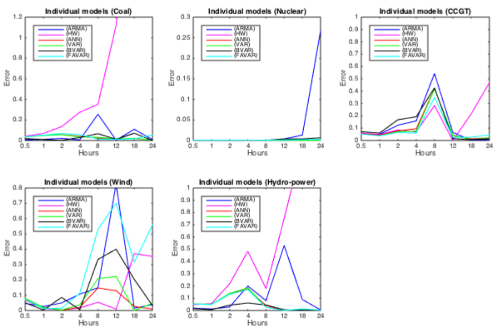
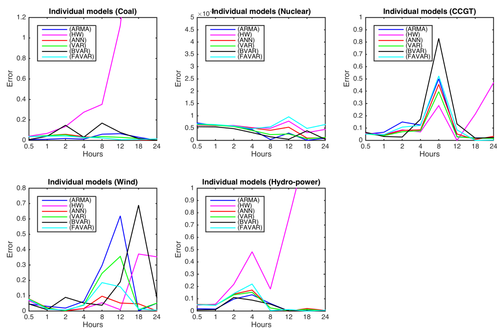


Figure 6: MSFE of IMS and DMS

The figure compares IMS and DMS forecasts precision for six classes of forecast models. The y axis is the value of MSFE, and the x axis is the forward prediction steps.



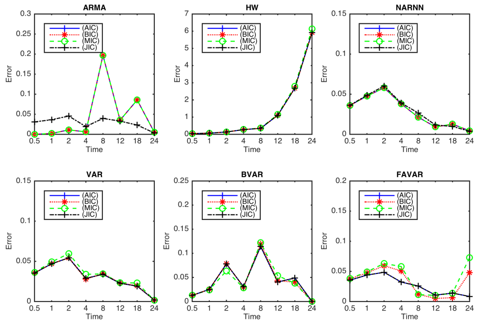
(a) IMS



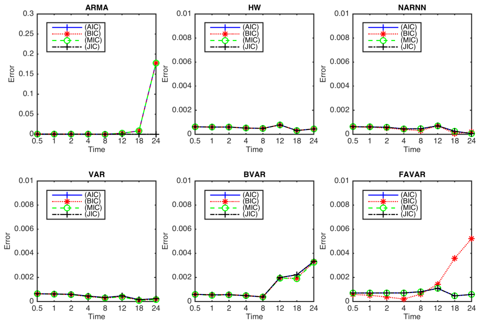
(b) DMS

Figure 7: Comparison among Information Criteria

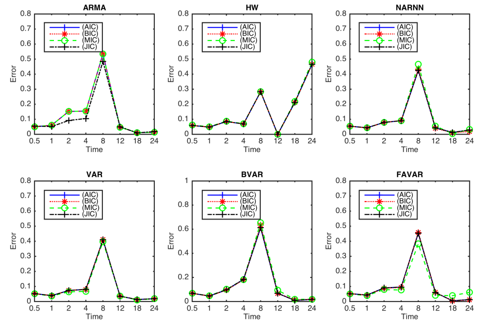
The figure compares model selection information criteria for six classes of forecast models. The y axis is the value of MSFE, and the x axis is the forward prediction steps.



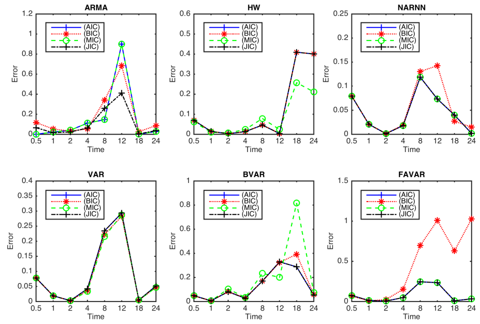
(a) Coal



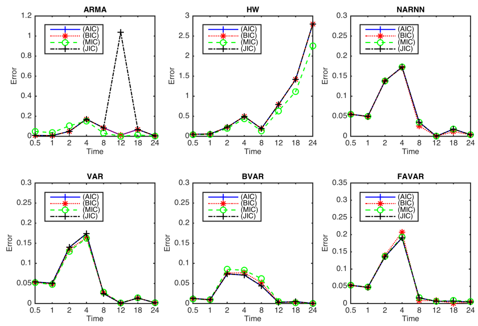
(b) Nuclear



(c) CCGT



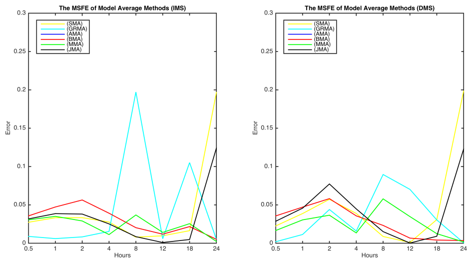
(d) Wind



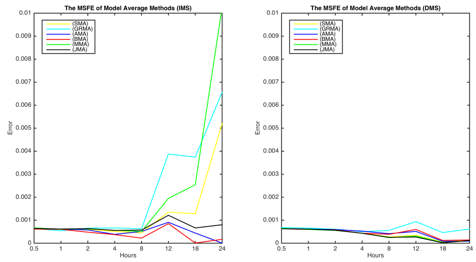
(e) Hydro-power

Figure 8: Comparison among Model Averaging Methods

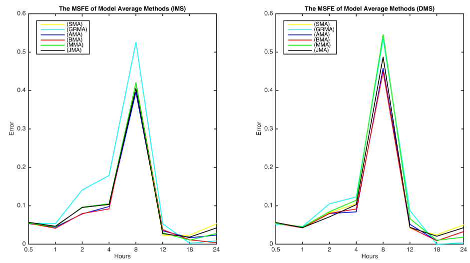
The figure compares six types of forecast model averaging methods within six types of forecast models. The y axis is the value of MSFE, and the x axis is the forward prediction steps.



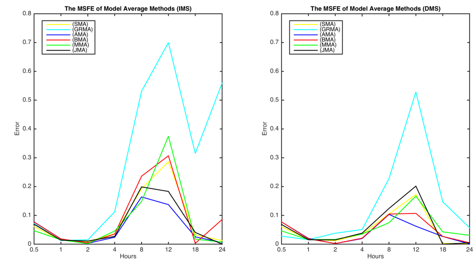
(a) Coal



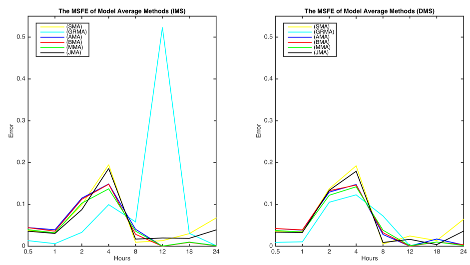
(b) Nuclear



(c) CCGT



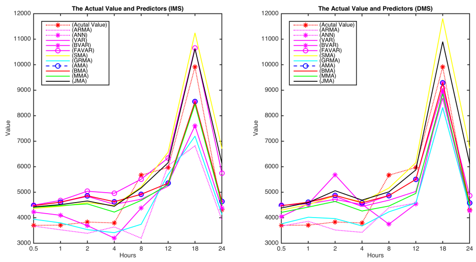
(d) Wind



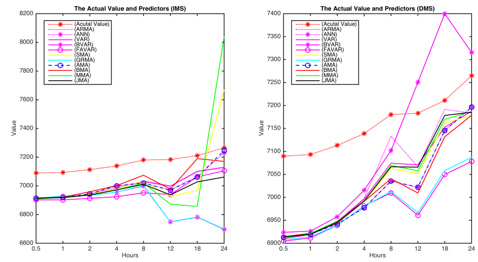
(e) Hydro-power

Figure 9: The Model Averaging vs Benchmark Models

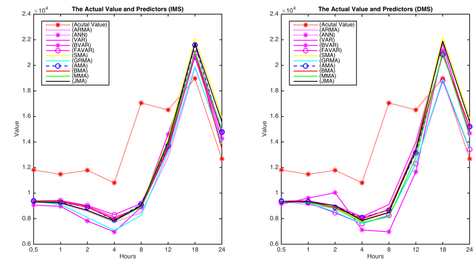
The figure reports six types of model averaging forecasts from six sets of forecast models. The y axis is the value, and the x axis is the forward prediction steps. We do not plot the results of HW model considering its poor performances.



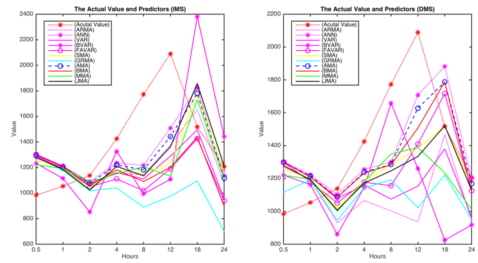
(a) Coal



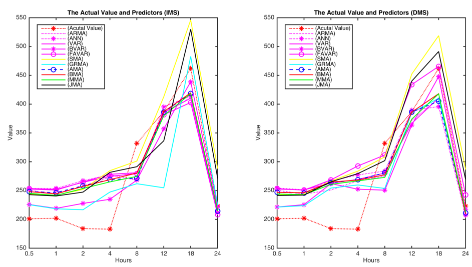
(b) Nuclear



(c) CCGT



(d) Wind



(e) Hydro-Power