Electrical Power and Energy Systems 95 (2018) 427–439

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A monthly electricity consumption forecasting method based on vector error correction model and self-adaptive screening method

Hongye Guo [a], Qixin Chen [a,*], Qing Xia [a], Chongqing Kang [a], Xian Zhang [b]

[a] State Key Lab. of Power System, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China
[b] Power Exchange Center, State Grid of China, Beijing 100031, China

ARTICLE INFO

Article history:
Received 2 March 2017
Received in revised form 7 August 2017
Accepted 4 September 2017

Keywords:
Electricity consumption forecasting
External economic factors
Vector error correction model
Self-adaptive screening
Seasonal decomposition

ABSTRACT

Economic growth has greatly fluctuated around the world in recent years, and external economic factors (EEFs) have imposed more obvious effects on electricity consumption. To improve the accuracy and applicability of mid-term, especially monthly, electricity consumption forecasting, a novel monthly electricity consumption forecasting framework (denoted as SAS-SVECM for short) based on vector error correction model (VECM) and self-adaptive screening (SAS) method is proposed in this paper, which fully explores and integrates the potential impacts from and relationships between EEFs. The SAS-SVECM firstly implements X-12-ARIMA to extract seasonal peaks from the electricity consumption and EEF time series. Second, a VECM is used to address correlations and time lag effects between electricity consumption and EEFs. And a SAS method is proposed to identify the most possible influential EEF self-adaptively, which appropriately addresses the contradiction between data quantity and data length. The SAS-SVECM achieves significant forecasting accuracy enhancement and good adaptability. Finally, an empirical example, using real monthly electricity consumption and macroeconomic data of China (2000–2014), was studied to verify the effectiveness of SAS-SVECM.

1. Introduction

1.1. Background of mid-term electricity consumption forecasting

Electricity consumption forecasting is a significant issue in the field of energy economics, which can be classified into different categories. According to the determinacy of outputs, it is classified as deterministic forecasting [1–3] and probabilistic forecasting [4–6]. According to relative time scales, such as long-term, mid-term and short-term. Mid-term electricity consumption forecasting always refers to the monthly time scale, which is important in many business and government decision-making processes. On the one hand, economic statistics and government policies, which are closely related to electricity consumption forecasting, are always performed and analyzed by the month. On the other hand, decisions on equipment maintenance, fuel trading and bilateral transactions of electricity are generally performed months ahead. These decisions could be greatly affected by mid-term electricity consumption forecasting [7].

As a fundamental industry of the national economy system, the monthly forecasting of electricity consumption is inevitably influenced by external economic factors (EEFs) [8,9], which, in the past, were usually represented by several common macroeconomic indexes.

In recent years, economic growth has greatly fluctuated around the world, mostly slowing down in many countries and regions. Facing new situations, the conventional mid-term electricity consumption forecasting methods can no longer precisely characterize the influence of EEFs [10,11].

EEFs include detailed industrial data, commodity trade data, and price index data, with as many as hundreds of types, released on a monthly or seasonal basis. Theoretically, with sufficient EEF data, the influential factors on electricity consumption can be clearly identified, and, therefore, prediction will become more accurate. However, conventional electricity consumption forecasting models are unable to process a large number of EEFs. First, too many input data would introduce the "curse of dimensionality" into the forecasting models. Second, although there are many types of EEFs, the lengths of most EEF data are relatively short, considering the monthly or seasonally released frequency. It is difficult to solve for the parameters of the forecasting model by simply taking large amounts of EEFs as inputs without sufficient data length.

Currently, some papers have performed research on mid-term electricity consumption forecasting methods considering various EEFs. The existing forecasting methods can be generally classified

* Corresponding author.
E-mail address: qxchen@tsinghua.edu.cn (Q. Chen).

http://dx.doi.org/10.1016/j.ijepes.2017.09.011
0142-0615/

Nomenclature

Abbreviations

EEF	external economic factor
SVM	support vector machine
VECM	vector error correction model
SVECM	seasonal VECM forecasting
SAS	self-adaptive screening method
ADF	augmented Dickey–Fuller test
TC	trend components
SC	seasonal components
HC	holiday components
IC	irregular components
IFG	input factor group
VAR	vector auto-regression model
AIC	Akaike information criterion
SBC	Schwartz–Bayes criterion
OLS	ordinary least squares method
MAPE	mean absolute percentage error

Indices and sets

t	index of time periods
i	index of EEFs

Variables

l	the data length of input EEF
m	the number of input factors of VECM
p	the time lag period of VECM
q	the training month of self-adaptive screening
C_t	the multiplication of TC and IC of EEF
D_t	the training forecasting error parameter of the tth month
E_i	the actual electricity consumption
I_t	the irregular component in month t
H_t	the holiday component in month t
S_t	the seasonal component in month t
T_t	the trend component in month t
Y_t	the value of selected EEF in month t
$f_{i,t}$	the ith EEF's time series
$\boldsymbol{A}_k$	an $n \times n$-order matrix of VAR model
$\boldsymbol{F}_t$	the time series vector of EEF in month t
$\boldsymbol{U}_t$	the disturbance vector with n orders in VAR model
$\boldsymbol{\Pi}$	the influence matrix of VEC model
$\boldsymbol{\Gamma}_k$	the coefficient matrix of VEC model

into two categories. The first category focuses on statistical techniques. The major feature is to consider the relevancy of electricity consumption and EEFs and to build parametric models, which include the time series method [12–15] and the multivariate regression method [1,16–20]. The second category focuses on intelligent algorithms, which include artificial neural networks [3,21–23], support vector machines (SVM) [2,24–26] and so on [27]. These algorithms consider the past electricity consumption and EEFs as inputs and current electricity consumption as outputs.

More specifically, the first category analyzes the relationships between electricity consumption and EEFs using multivariable regression or correlation analyses. Because the impact of some EEFs on electricity consumption from might last for months, a precise analysis of the relationship between electricity consumption and EEFs needs to consider both all EEFs at the same time and trends implicit in the historical data. However, most of the existing parametric forecasting methods did not pay sufficient attention to time lag effects of EEFs. Forecasting is mostly performed based on the synchronous data of EEFs [1,16]. Although some researchers have considered time lag effects of EEFs, a determined period is always subjectively given to represent the time lag effects before model training [3]. Moreover, the quantity of the EEF considered in parametric methods is usually constrained by the chosen methods. Considering the hundreds of input data types, conventional parametric forecasting methods combining several regression functions do not have the ability to self-adaptively choose appropriate data as inputs. Some of papers have attempted to make a "filtrate" job in advance and then selected a small number of appropriate EEFs for electricity consumption forecasting [2,17]. This measure indicates that these forecasting methods have limitation on handling so large economic datasets.

The second category, which is mostly based on AI techniques, uses machine learning or pattern recognition abilities to organize the input factors adaptively. However, one weakness of AI techniques is that the entire computational process is mostly a "black box", which is not as understandable as the parametric methods [28]. Therefore, it is extremely difficult to understand the process and the interaction of how inputs influence the outputs. From an economic perspective, it would be important to understand the relationships between mid-term electricity consumption and the corresponding EEF. Moreover, the application of AI techniques involves a large number of training data. Even though the available data volume is increasing rapidly, considering the hundreds of factor types, it would still take a long time for the requirements of conventional AI methods' training data amount to be met.

In summary, because the set of EEFs has greatly increased, conventional parametric models might be too simple to capture the abundant information hidden within the data, and AI models do not solve the aforementioned contradiction between data quantity and data length. Therefore, a new monthly electricity consumption forecasting method that can effectively and credibly integrate large amounts of EEF data and thus self-adaptively improve the forecasting accuracy is needed.

1.2. Used and proposed forecasting methods

To handle the contradiction between data quantity and data length, the vector error correction model (VECM) [29] has been broadly applied in macro-econometrics in recent years [10,30]. Several studies have used VECM to describe the relationship between electricity consumption and EEFs [10,11,30,31] and performed forecasting of electricity consumption. However, the types of input data for these works were limited, and the forecasting periods were usually long, such as annually or quarterly. VECM is essentially a type of parameter model; thus, the data length requirements are relaxed relative to the commonly used AI models. By building up the influence matrix and the coefficient matrixes, VECM is able to represent the long-run (historical trends) and the short-run (recent changes) equilibrium relationships between input factors and identify the time lag effects in the most relevant sense. As a fundamental industry of the national economy system, electricity consumption can be integrated as a regular factor in the VECM; thus, hidden correlated information between electricity consumption and EEFs can be clearly identified and integrated into