Applied Energy 156 (2015) 502–518

Contents lists available at ScienceDirect

Applied Energy

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

Multivariate statistical and similarity measure based semiparametric modeling of the probability distribution: A novel approach to the case study of mid-long term electricity consumption forecasting in China

Zhen Shao [a,b,*], Fei Gao [a], Qiang Zhang [a], Shan-Lin Yang [a,b]

[a] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[b] Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui 230009, PR China

HIGHLIGHTS

- A novel semiparametric approach is proposed for electricity demand density forecast.
- A new similarity measure strategy is designed based on K–L divergence analysis.
- We consider both climate exclude and include conditions for mid-long term forecast.
- High predictive accuracy is demonstrated when using the finance related variables.
- We analyze possible future demand for China's energy consumption system until 2020.

ARTICLE INFO

Article history:
Received 7 January 2015
Received in revised form 22 June 2015
Accepted 14 July 2015

Keywords:
Semi-parametric regression
Similarity measure
Probability distribution forecast
Mid-long term demand forecast
Variable simulation

ABSTRACT

To achieve the goal of drawing up optimal plans for power generation, decision makers need an appropriate methodology to effectively identify the pivotal aspects of electricity consumption fluctuation and anticipate the future trend. The parameter identification of conventional statistical approach mainly relies on distributional assumptions and functional form restrictions, which might be problematic for the real application. This paper addresses these issues by implementing a novel semi-parametric modeling approach, which is suitable for investigating the uncertainties in the mid-long term forecast and estimating the probability distributions of future demand. To identify the significant impact factors of the electricity consumption, a new Kullback–Liebler (K–L) divergence based similarity measure strategy is designed. A case study concerning the electricity demand forecasting in China demonstrates the applicability of the proposed approach and verifies the feasibility of establishing explicit functional dependency between external variables and electricity consumption. Despite the complexity, notable reductions in the number of forecasting error are obtained due to the adoption of three indicators: deposits in financial institutions, exports, and imports.

1. Introduction

To achieve a desired trade-off between energy production capacity, energy efficiency and reliability, the deployment of smart grid is crucial. One of the primary tasks of smart grid is to address the current concerns in electricity systems, including managing the power generation and transmission, expanding the use of low-carbon technologies, etc. Regional power grid interconnection, long-distance power transmission and scheduling are immediate priorities to the development of smart grid in China. But, at the same time, precise electric planning and budgeting are also emerging issues.

As the world's largest consumer and producer of energy, China will reduce its carbon dioxide emissions per unit of GDP, by 17 percent of 2010 levels by 2015 [1]. At the end of 2015, China's renewable energy power (hydropower, wind energy, etc.) generation will accounts for more than 20% of the total generating capacity [2]. Despite all of this, it is important to note that the energy resources which scattered widely across China are not evenly distributed. China's electric power must be transported over long distances from the west to the southeast. In this regard,

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China. Tel.: +86 055162904935; fax: +86 05512905263.

E-mail address: shaozhen@mail.hfut.edu.cn (Z. Shao).

http://dx.doi.org/10.1016/j.apenergy.2015.07.037
0306-2619/

Nomenclature

Y_i	electricity demand, Eqs. (1) and (2)
X_i	vector of explanatory variables for the ith of n observations, Eqs. (1) and (2)
β_i	unknown parameter vector to be estimated, Eqs. (1) and (2)
$g(t_i)$	unknown smooth function (nonparametric portion of the model) to be estimated, Eqs. (1) and (2)
ε_i	residual which has the properties $E(\varepsilon_i) = 0$, $E(\varepsilon_i\varepsilon_j) = 0$, $(i \neq j)$, $\mathrm{Var}(\varepsilon_i) = \sigma^2$, Eqs. (1) and (2)
B	backward shift operator, Eq. (2)
ω_i	the d-th difference of the series $g(.)$, Eq. (2)
γ_i	sequence of independent normal deviates with common variance σ^2, Eq. (2)
φ_i	parameter of the autoregressive part, Eq. (2)
θ_i	parameters of the moving average part, Eq. (2)
∇^d	backward difference, Eq. (2)
p	order of the autoregressive part, Eq. (2)
q	order of the moving-average part, Eq. (2)
d	degree of differencing, Eq. (2)
L_2^h	Hilbert space of measurable H-square integrable functions
$\hat{\beta}$	estimate of β, Eq. (3)
$\hat{g}(t)$	estimate of $g(t)$, Eq. (3)
w_i	weights of the penalized spline estimation, Eq. (3)
λ	penalty parameter makes balance between prediction accuracy and sparsity, Eq. (3)
$J(g)$	nonparametric function quantifies the roughness of g, Eq. (3)
g^m	m order fractional derivative of unknown function $g(.)$, Eq. (4)
Q	$n \times (n-2)$ matrix on the given knot sequence (t_1, t_n), Eq. (5)
T	$(n-2) \times (n-2)$ strictly diagonal dominant matrix on the given knot sequence (t_1, t_n), Eq. (5)
N	$n \times q_1$ incidence matrix, where $q_1 \geqslant 2$, Eqs. (6) and (7)
W	$n\times n$ diagonal matrix of weights w_i, Eq. (6) and (7)
α	the degree of similarity measure, Eq. (8)
D_i^y	represents the distance between test sample and all the feature components of the original sequence, Eq. (8)
D_i^{-y}	represents the distance between test sample and all the samples that belongs to other classes (illusive components of original sequence), Eq. (8)
H_n	n-simplex, Eq. (9)
$p(x)$	probability mass function, Eq. (10)
$q(x)$	probability mass function, Eq. (10)
$KL(\cdot\|\|\cdot)$	K–L divergence rate, Eqs. (10) and (11)
M^∞	infinite cartesian product
$\tilde{P}$	probability laws on the infinite cartesian product M^∞, Eq. (11)
$\tilde{Q}$	probability laws on the infinite cartesian product M^∞, Eq. (11)
$\tilde{P}_l$	the marginal distributions of $\tilde{P}$, Eq. (11)
$\tilde{Q}_l$	the marginal distributions of $\tilde{Q}$, Eq. (11)
$R(\tilde{P}\|\|\tilde{Q})$	a generalization of the K–L divergence rate, Eq. (11)

Acronyms

KL	Kullback–Liebler
SPR	semi-parametric regression
GCV	generalized cross-validation
KLDSM	Kullback Leibler divergence-based similarity measure
EEMD	Ensemble empirical mode decomposition
IMF	intrinsic mode functions
TRSCG	total retail sales of consumer goods
WRTTMS	wholesale and retail trades total merchandise sales
TIE	total import and export
CPI	consumer price index
RPI	retail price index
PPI	producer price index
RMFP-PPI	purchasing price indices of raw material, fuel and power
EPPI	electricity producer price index
MPMI	manufacturing purchasing managers index
SR	state revenue
M2	money and quasi-money
SD	savings deposit
DFI	deposits in financial institutions
MAT	monthly average temperature
MAP	monthly average precipitation
FA	Factor analysis
MSPR	multivariate semi-parametric regression
MAPE	mean absolute percentage error

electric power transportation and scheduling present challenges for planners. Specifically, economic growth together with other factors constitute the distinct growth phase of China's electricity consumption. As is illustrated in Fig. 1, industrial and residential consumption together account for nearly 90% of the total electricity consumption in China. Consumption by other sectors is of minor importance. Therefore, the precise estimation of electricity consumption is of fundamental importance to keep the national economy running smoothly. Considering the case of electricity demand planning, decision makers must decide how to allocate power resources to balance the brisk demand with the finite resource.

Mid-long term demand forecasting plays a critical role in monitoring and planning the transportation of electric power. As in the case of electricity demand analysis, the aim of mid-long term load forecasting is to obtain the best possible estimate of the future electricity demand over a long time horizon (typically for prediction range that is longer than one year). With the fast advance in sensing measurement technology and computing technology, smart grid has become an indispensable component of modern society. These advances will most likely lead to an adjustment of the core mission of traditional electricity demand forecasting, for instance, moving from the traditional way of utilizing single historical data to the multi-parameter optimization of considering more complex affecting factors.

By developing a novel approach for combining semiparametric model, bootstrap resampling based estimates, and K–L divergence-based similarity measure, this paper makes three key contributions to the solution of mid-long term electricity demand forecasting in China: (1) A semi-parametric regression model composed of time series smoothing is proposed for mid-term load forecasting, which does not impose any constraints on the assumption of the model distribution and enables the uncertainty in the measurement of external indicators to be considered in an unspecified functional form (nonparametric portion of the model). (2) The penalized smoothing based bootstrap estimate is developed to provide a robust estimation of parameters and generate density forecasts of the future demand. After obtaining probability density forecasts, it is more likely to gain a deeper insight (compared with the point forecasts) into the future fluctuation trend of electricity