Can Internet Search Behavior Help to Forecast the Macro Economy?
Taoxiong Liu Xiaofei Xu Tsinghua University

INTRODUCTION
In macroeconomy analysis, two types of data can be applied, namely the structured and unstructured data. Government statistics are well structured while Internet search behavior information belongs to non-structured data. This research uses 6 types of models to forecast the macroeconomic aggregate. By comparing different models, the optimal forecast model is selected. We shows that the Internet searching behavior can help forecast the macro economy. Moreover we find that the correct way for variables selection with structured and unstructured data is the two-step method. Firstly only the government statistics are used and the conditional optimal model is selected. Secondly, the internet search data are added into these model and the optimal model is determined.

METHOD
a. Constructs 6 types of models for the forecast of quarterly GDP - Use the explained variable information alone
- Separately use the government statistical information and Internet searching behavior information
- On the basis of explained variable information, the government statistical data and Internet searching behavior data are added respectively
- Use the explained variable information, government statistics data and Internet searching behavior information at the same time
b. Model selection criterion: according to the BIC value (or training MSE)
c. A comparative analysis
- By using the automatic model selection in Oxmetrics software, we find its results are worse than ours. This further illustrates the effectiveness of the model selection and forecast method of this study.

Empirical result
a. Regression and prediction by Auto-regressive model of GDP. According to the BIC, only GDP4 is selected as explanatory variable.

Table 1 GDP autoregression results

Table 2 Government statistics and Internet searching behavior

Table 3 GDP lags plus government statistics and GDP lags plus Internet searching behavior

Table 4 Adding Search Behavior into the Model without Search Behavior

Conclusion
In macroeconomic forecast, two types of data might be involved, structured data and unstructured data. We use government statistics and Internet search behavior information to forecast GDP. By comparing different models, the optimal model is selected. We shows that the Internet searching behavior can help forecast the macro economy. Moreover we find that the correct way for variables selection with structured and unstructured data is the two-step method. Firstly only the government statistics are included and the conditional optimal model is selected. Secondly, the internet search data are added into the model and the optimal model is determined.

by adding unstructured data (Zs) to the conditionally optimal model within the models having GDP lags and two other variable of structured data (Xs), we find that the forecast can still be improved. (Two-Step Method)