

The Least Energy Criterion *LEC* and Its Application in Model Selection

Mu Qiao^{1,a}, Peng Wang^{2,b}

¹School of Mathematics, Jilin University, Changchun, 130012, China

²School of Management, Jilin University, Changchun, 130025, China

^aqiaomu12@126.com, ^bwangpeng19860123@163.com

Keywords: least energy principle; information entropy; model selection

Abstract. This paper considers the statistic *LEC* and its application on econometric model selection. The idea is to apply the concept of least energy principle, and based on information entropy propose the least energy criterion expression *LEC*. It is shown that using the idea of comparing the *LEC* of model residual sequences and selecting the least *LEC* one as the real model is feasible. For practice, this paper takes the data of China export volume from 1978 to 2011 as an example to verify the idea. The conclusion shows that using least energy criterion to select model is effective.

1. Introduction

Model selection is an important part of modeling econometrics time series. If there are several sufficient models for a certain data set, the model selection method could help us choose the best one. In 1973, Akaike generalized the criterion to ARMA model order estimation, which is based on model forecast error to estimate the order of autoregression model. It is a least information criterion which called AIC Akaike Information Criterion [1]. After that in 1976, Shibata proved that AIC may overestimate the autoregression order, and then Akaike proposed the Bayes' generalization of AIC, which is BIC criterion [2]. Parzen aimed at AR model proposed autoregression transfer function criterion CAT in 1977 [3]. Based on residual sum of squares criterion, Mallows proposed C_p statistic criterion which is similar as AIC [4]. Each of these criteria has their advantages and application conditions. This paper considers using the idea of least energy principle and proposing a model selection method based on residual analysis.

The least energy principle is one of the most basic principles of our nature world. It is a principle that a system would always adjust itself until the least gross energy condition, in order to make itself a stable and equilibrium state, this principle is suitable for both of the macro and micro world. It is not only a physical principle, but also can be found similar phenomenon in economic field. For example, an isolated economic system would constantly adjust itself to an equilibrium state, and then stop developing; the energy of this system is also the lowest by that time.

As is known to all of the distribution functions, normal distribution is the most widely used one. Writer believes normal distribution, relatively, has the lowest energy, and it is an embodiment of the least energy principle in Statistics. For example, the familiar time series central limit theorem. Hamilton et al. [5-8] wrote in their literature that with the increase of sample size T , the probability density of the sample mean \bar{Y}_T for an *i.i.d.* sequence collapsing towards a point mass at μ , and the sequence $\sqrt{T}(\bar{Y}_T - \mu)$ converges in distribution to a normal random variable, that is if Y_t is *i.i.d.* with mean μ and variance σ^2 , then

$$\sqrt{T}(\bar{Y}_T - \mu) \xrightarrow{L} N(0, \sigma^2)$$

(1)

The central limit theorem can be also interpreted as that with the increase of sample size T , the sequence $\sqrt{T}(\bar{Y}_T - \mu)$ converges to the least energy distribution.

2. The least energy criterion *LEC*

First of all, imaging that a distribution of random variable is an energy field, and the expectation

is the barycentre of the field, every random variable is a mass point fall into the field. If the energy of the field is zero, then each mass point should fall onto the same position, that is the expectation. Nevertheless, the energy of the field makes them deviate from the expectation. Higher energy could make them deviate farther, however, the sum of the energy of each RV mass point it carries is exactly the gross energy of the distribution field. It can be simply found from our conjecture that the energy of a distribution is directly proportional to the variance σ^2 .

The actual thing that data provides is information; therefore, the energy of a sequence could be named as information energy. Information is a very abstract concept, which is hard to be quantified. Fortunately, Shannon (1948) proposed the concept of information entropy, which is an abstract mathematic concept [9]. Its expression is defined as

$$H(x) = H(P_1, P_2, \dots, P_n) = -\sum_i P(x_i) \log P(x_i). \quad (2)$$

Information entropy could be interpreted as the appearance probability of a sort of specific information. The basic effect of information is to eliminate indeterminacy. If a variable has high information entropy, that is it has higher uncertainty. In other words, an ordered system has lower information entropy and, on the other side, a disordered system has higher information entropy. Therefore, so to speak, information entropy is used for measuring the degree of ordering of a system. It has been widely applied in econometric fields [10-12]. As we know when fix mean μ and variance σ^2 , the normal distribution has the highest information entropy, it is a kind of 'natural selection'. Normal distribution is, relatively, also the least energy one, so we can found the energy is inversely proportional to the information entropy.

Based on the concept of information entropy and above analysis we propose the expression of the energy of a distribution.

Definition 1. *If a distribution has variance $\sigma^2 \neq 0$, and the variable is continuous, then we define LEC which*

$$LEC = \frac{\sigma^2}{-\int p(x_i) \ln p(x_i)} \quad (3)$$

as the energy of the distribution. If the variance $\sigma^2 = 0$, the $LEC=0$.

Definition 2. *If a distribution has variance $\sigma^2 \neq 0$, and the variable is discrete, then we define LEC which*

$$LEC = \frac{\sigma^2}{-\sum p(x_i) \ln p(x_i)} \quad (4)$$

as the energy of the distribution. If the variance $\sigma^2 = 0$, the $LEC=0$.

We can easily calculate the energy statistic LEC for several representative distributions should be as following,

1. For the Normal distribution $N(\mu, \sigma^2)$,

$$LEC_{norm} = \frac{\sigma^2}{\ln(\sigma\sqrt{2\pi e})}. \quad (5)$$

2. For the Gamma distribution $\Gamma(\alpha, \beta)$,

$$LEC_{gamma} = \frac{\alpha\beta^2}{\alpha + \ln \beta + \ln \Gamma(\alpha) + (1 - \alpha)\psi(\beta)}. \quad (6)$$

3. For the Exponential distribution $Exp(\lambda)$,

$$LEC_{exp} = \frac{\lambda^{-2}}{1 - \ln \lambda}. \quad (7)$$

From the definition we can also found that the statistic LEC should has natures as following,

1. If fix variance σ^2 , where $\sigma^2 \neq 0$, then normal distribution has the least LEC of all distributions.
2. The LEC of normal distribution would increase with the variance σ^2 increasing.
3. The least LEC is 0 when $\sigma^2 = 0$.

Definition 3. Let $X_t = (x_1, x_2, \dots, x_T)$ are T sample observations of an i.i.d. sequence, then the energy of the sequence is defined as

$$LEC = \frac{s^2}{-\sum_{i=1}^T p(x_i) \ln p(x_i)}, \quad (8)$$

where LEC expresses the energy, s^2 is the sample variance, and $p(x_i)$ expresses the probability of that x_i take a certain value or fall into a certain interval.

The statistic LEC of a sequence describes the energy for transferring into information of the sequence. Therefore, if the LEC of the residual sequence of a model is very small, that is the residual sequence has very low energy to transferring into useful information, then we could say that the model is better. In other words, the model which has lower LEC residual sequence could explain data generating process better. Writer defined this model selection method as the Least Energy Criterion.

Owing to the difference of unit and dimension, the statistic LEC is a relative quantity, which can be only compared in identical system, but not suitable for the measurement between tow or more different systems.

3. Application

In the financial and economic sphere, the AR model is more widely used than ARMA and MA, due to the parameter identification of AR is simpler and it has well instantaneity. Moreover, it has been proved that ARMA and MA model could be approximated by the AR model.

Lots of current macro economic data or its logarithmic form obeys AR(1) process and most of them have a unit root. We can use the method that comparing the LEC of residual sequence to test AR(1) unit root. In other words, respectively using AR(1) models with unit root and without unit root to fitting data generating process. If the residual sequence of the model with unit root has the least LEC , then we can estimate that the data generating process is a unit root process, it is because residual sequence has lower LEC means the model could explain the data better.

For empirical study, this paper takes the logarithmic sequence $\ln EX_t$ of China export volume from 1978 to 2011 as example to validate the idea. Data show as in Figure 1, and we analyze the sequence in steps as follow.

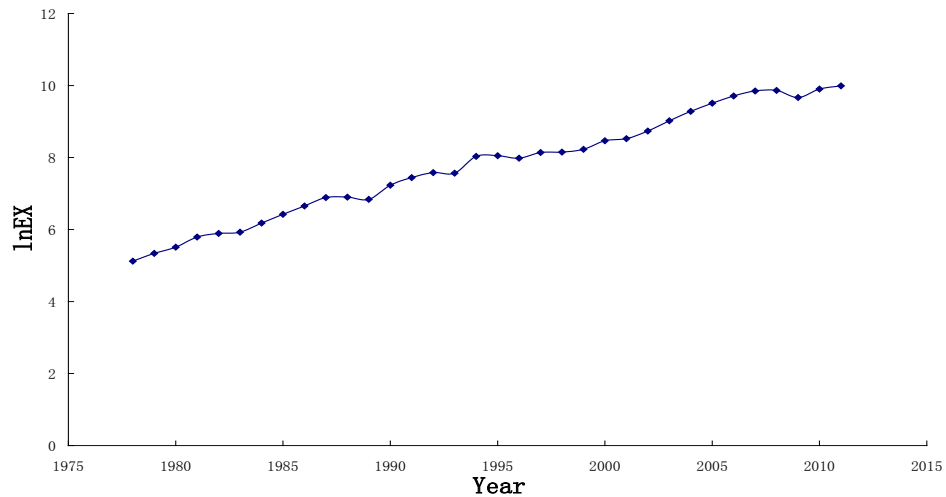


Figure 1. 1978-2011 China export volume

Firstly, according to the preliminary judgment, we use the AR(1) model to fitting the data generating process,

$$\ln EX_{t+1} = \beta \ln EX_t + \varepsilon_t, \quad (9)$$

$$\text{or } \ln EX_{t+1} - \mu = \beta \ln EX_t + \varepsilon_t, \quad (10)$$

in this paper we respectively let $\beta=1, 0.95, 0.9, 0.8, 0.6$ five values and use these five models to fitting data, then obtain five residual sequences.

Secondly, define the probability $p(x_i)$ of the data in these residual sequences. First of all, sorting the residual sequence by size, and then divide the interval into groups in average between the maximum and minimum value, where the number of groups is calculated by the Sturges formula:

$$k=1 + 3.3\ln 34 \approx 6, \quad (11)$$

And then calculate the ratio of the data size of a certain group and the sample size T , it is the probability $p(x_i)$ of all the data which fall into this group.

Finally, calculate the *LEC* of the five residual sequences respectively and compare the size of them. The calculating result show in Table 1,

Table 1. The calculating result

β	0.6	0.8	0.9	0.95	1
<i>LE</i>	0.032472	0.008636	0.003162	0.001959	0.001904

We can see from Table 1 when $\beta=1$, that is the model with unit root has the least residual energy *LEC* 0.001904. It means we can come to a conclusion that the data generating process of the sequence $\ln EX_t$ is a unit root process and the export volume of China is unstable.

In order to prove the result is correct we use the traditional unit root test method ADF on the same data, the result shows in Table 2,

Table 2. The result of ADF test

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.569868	0.2954
Test critical values:		
1% level	-4.262735	
5% level	-3.552973	
10% level	-3.209642	

Form Table 2 we can see that no matter choosing critical value 1%, 5% or 10%, we can not reject

the null hypothesis that the data generating process has a unit root. The result of ADF test identifies with that of the least energy criterion; therefore, we can draw a conclusion that using least energy criterion to select model is effective.

4. Conclusion

Compared with the traditional model selection methods, least energy criterion has its own advantages. Firstly, most of the methods need a hypothesis that the type of distribution is known, but the LEC does not. Secondly, the least energy criterion only need to calculate the LEC of residual sequence, it is much simpler in calculation than other methods. Least energy method also has its restriction, such as before analysis we need to estimate the model type according to the data. Writer believes the statistic *LEC* would have a wider range of application in other fields and it will leave to the future studies.

Reference

- [1] Akaike H: Fitting Autoregressive for Prediction Models, *Statist Math*, Vol.21 (1969), p.243-247.
- [2] Akaike H: *Information Theory and an Extension of the Maximum Likelihood Principle*, in *Petrov and Czaki Eds*, Proceedings of the 2nd International Symposium on Information Theory, (1976), p.267-281.
- [3] Parzen E: *Multiple Time Series Determing the Order of Approximating Autoregressive Schemes*, (Multivariate Analysis IV, North-Holland, 1977).
- [4] Mallows C.L: *Choosing Variables in a linear Regression*, Presented at the Central Regional Meeting of the IMA, Manhan, Kansas, (1964), p.178-180.
- [5] James D. Hamilton: *Time Series Analysis*, (Princeton: Princeton University Press, 1994).
- [6] Donald W. K. Andrews: Laws of large Numbers for Dependent Non-Identically Distribution Random Variables, *Econometric Theory*, Vol. 4, No.3 (1988), p. 458-467.
- [7] Peter C. B., Phillips and Victor Solo: Asymptotics for Linear Processes, *The Annals of Statistics*, Vol. 20, No.2 (1992), p. 971-1001.
- [8] Halbert White: *Asymptotic Theory for Econometricians*, (Orlando, Fla.: Academic Press, 1984).
- [9] C. E. Shannon: A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol. 27 (1948), pp. 379-423.
- [10] Amos Golan: Information and Entropy Econometrics- Editor's View, *Journal of Econometrics*, Vol.107 (2002), p.1-15.
- [11] Alastair R. Hall, Atsushi Inoue, Kalidas Jana, Changmock Shin: Information in generalized method of moments estimation and entropy-based moment selection, *Journal of Econometrics*, Vol.138 (2007), p.488-512.
- [12] Y. H. Sun, H. B. Zhou: Measurement Uncertainty Estimation Method based on Maximum Entropy, *Statistics and Decision*, Vol. 12 (2008), p. 141-144.