On the Model Selection Criteria for Demand System: Theil's Minimum Entropy Measure and its Modification with Resampling Method

Mototsugu FUKUSHIGE and Kosuke OYA,
Department of economics, Kobe University and Graduate School of Economics, Osaka University

ABSTRACT: Several models are proposed for analyzing consumer demand system. Empirical researchers usually, adopting one or more models and testing some constraints based on the demand theory (positivity, homogeneity, symmetry), evaluate the plausibility or superiority of the models that they adopted. There are some reasons why they rarely adopt a statistical model selection procedure in their evaluation. First of all, dependent variables, which represent consumer demand, are different in each model respectively, so it is difficult to compare their likelihood based upon some specific data generating process. Another reason is originated from the properties of the share functions. Adding-up condition of all the item shares makes the sum of the disturbances must be zero. In other words, the joint distribution of the disturbances must be degenerated. This property forces the researcher to drop one of the share equations in practical estimation. Then, the number of estimated equations of the share model and of the level model are different. This is another source that makes likelihood comparison difficult. Theil's minimum entropy measure, which is calculated from the estimated and actual shares and does not depend on the number of the estimated equations, is a candidate for a practical model selection criterion, but its empirical performance is not studied carefully and sufficiently. In this paper, we conduct a Monte Carlo simulation. Moreover, utilizing the Bootstrap methods, we modify this measure more robust criteria. A simulation result with Japanese households' consumption data shows that Theil's measure and its modified version with Bootstrap method performs well.

1. INTRODUCTION

Several different types of equation systems are estimated for analysis of consumer demand. Each equation system is based on the different economic theory that has been exclusively established. We may decide which economic theory is statistically supported by actual data using standard statistical procedures such as model selection by information criterion and hypothesis testing procedure. In the case of consumer demand analysis, however, we can not use of standard procedures such as R^2, *likelihood* and *information criterion based on likelihood* since the different dependent variables are used in the different models of consumer demand systems being compared and selected.

The objective of this paper is to establish the model selection procedure for the consumer demand systems. The plan of this paper as follows. The theoretical models and related theoretical constraints of demand analysis that should be satisfied are presented in the next section. The model selection procedure is proposed in section 3. The proposed procedure is based on the minimum entropy measure using the estimated budget share. A series of Monte Carlo is conducted to see the performance of proposed procedure in section 4. Some concluding remarks are in last section.

2. DEMAND SYSTEM AND CONSTRAINTS

We will present the typical models of consumer demand system and theoretical constraints called as 'Law of Demand'.

2.1 Rotterdam Model

First model is Rotterdam model:

$$\overline{w}_{it} Dq_{it} = a_i + b_i \sum_{k=1}^{N} \overline{w}_{kt} Dq_{kt} + \sum_{k=1}^{N} c_{ik} Dp_{kt} \quad (1)$$

where

$$Dq_{it} = \log q_{it} - \log q_{it-1} \quad (2)$$

$$Dp_{it} = \log p_{it} - \log p_{it-1} \quad (3)$$

$$\overline{w}_{it} = \frac{1}{2}(w_{it} + w_{it-1}) . \quad (4)$$

w_{it} is the observed budget share of the i-th category in the t-th period, p_{it} and q_{it} are the price index and the real expenditure for the i-th good, in t-th period, respectively.

Theoretical constraints, we have to impose and check whether they are consistent with actual data or not, are the *adding-up, homogeneity, symmetry* and *negativity* conditions. These conditions can be represented using the following parameter constraints as:

(i) *adding-up:*

$$\sum_{k=1}^{N} a_k = 1,\ \sum_{k=1}^{N} b_k = 0 \text{ and } \sum_{k=1}^{N} c_{kj} = 0 \qquad (5)$$

(ii) *homogeneity:*

$$\sum_{k=1}^{N} c_{jk} = 0 \qquad (6)$$

(iii) *symmetry:*

$$c_{ij} = c_{ji} \qquad (7)$$

(iv) *negativity:*

$$C = [c_{ij}] \text{ is negative semi-definite} \qquad (8)$$

where the adding-up constraints is used for estimation, is not tested. The negativity constraint is not tested in this paper for the simplicity.

The constraints that should be tested here are the *homogeneity* and *symmetry*. Tests of these conditions are considered as tests of 'Law of Demand'. For more details, see Barten [1964].

2.2 Linear Expenditure Model

Adding-up, homogeneity and symmetry conditions have already been imposed in following linear expenditure system. This means that there is no testable constraint in this model.

$$p_{it} q_{it} = b_i p_{it} + a_i \left(X_t - \sum_{k=1}^{N} p_{kt} b_k \right) \qquad (9)$$

where X_t is the total expenditure in the t-th period. The constraint we impose for estimation of (9) is

$$\sum_{k=1}^{N} a_k = 1. \qquad (10)$$

This is not constraint for testing. See Stone [1954] for more details of this model.

2.3 Almost Ideal Demand System Model

Almost Ideal Demand System (AIDS) model is can be written as

$$w_{it} = a_i + b_i \log\left(\frac{X_t}{P_t^*} \right) + \sum_{k=1}^{N} c_{ik} \log p_{it} \qquad (11)$$

where

$$\log P_t^* = \sum_{k=1}^{N} w_{ik} \log p_{it} .$$

The adding-up constraints are

$$\sum_{k=1}^{N} a_k = 1,\ \sum_{k=1}^{N} b_k = 0 \text{ and } \sum_{k=1}^{N} c_{kj} = 0 . \qquad (12)$$

These are used for estimation and not for test. The homogeneity and the symmetry are

$$\sum_{k=1}^{N} c_{jk} = 0 \text{ and } c_{ij} = c_{ji} \qquad (13)$$

respectively.

These three models are candidates for our model selection of the consumer demand system. The parameters we have estimate are a_i, b_i, c_{ij} for i=1,2,...,N and j=1,2,...,T. N is the number of goods and T is the number of periods. See Deaton et al. [1980] for more details.

3. MODEL SELECTION PROCEDURE

We propose the model selection procedure for the models of consumer demand systems in this section. We will adopt two estimation procedures. One is the unconstrained and the other is the constrained estimation.

3.1 Estimation

Denoting the parameters θ, we describe all the models in the previous section as

$$y_{it}^{(m)} = g^{(m)}(Z_{it}^{(m)};\theta^{(m)}) + \varepsilon_{it}^{(m)} \qquad (14)$$

where the superscript (m) represent which model are estimated. (RT), (LE) and (AI) are Rotterdam model, Linear Expenditure model and Almost Ideal demand system model. In the case of (RT),

$$y_{it}^{(RT)} = \overline{w}_{it} Dq_{it}, \qquad (15)$$

$$Z_{it}^{(RT)} = \left(\sum_{k=1}^{N} \overline{w}_{kt} Dq_{kt} \quad Dp_{1t} \quad \cdots \quad Dp_{Nt} \right) \qquad (16)$$

and

$$\theta_i^{(RT)} = \left(a_i \quad b_i \quad c_{11} \quad \cdots \quad c_{1N} \right). \qquad (17)$$

Total number of equations in each system is N, however we estimate N-1 equations since we adopt the adding-up constraint for estimation. When we impose the additional constraints such as homogeneity and symmetry, we apply the constrained least squares estimation. We define the unconstrained and constrained estimators as $\hat{\theta}_i$ and $\tilde{\theta}_i$, respectively.

3.2 Criterion

Although we have different dependent variables in the different models, we can obtain the estimator of budget share that implied by each model. The budget share estimators based on the unconstrained and constrained estimator of θ can be represented as the function of $\hat{\theta}$ and $\tilde{\theta}$, *i.e.* $w_{it}(\hat{\theta})$ and $w_{it}(\tilde{\theta})$, respectively.

Next we construct the criterions for the model selection of the consumer demand system using the budget share estimators defined above.

The information measure known as *Average Information Inaccuracy*, e.g. Finke et al.[1984], is

$$\bar{I} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} w_{it} \log\frac{w_{it}}{\hat{w}_{it}} \tag{18}$$

where $\hat{w}_{it}$ is the forecast of the budget share. This measure evaluates the forecasts of the budget shares.

One simple model selection measure can be defined based on (18) with constrained estimator $\tilde{\theta}$ as

$$\tilde{I}^{(m)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} w_{it} \log\frac{w_{it}}{w_{it}(\tilde{\theta}^{(m)})}. \tag{19}$$

The constrained estimators of $\theta^{(m)}$ are used for construction of this measure (19). Alternatively, we can propose the measure using the unconstrained estimators as

$$\hat{I}^{(m)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} w_{it} \log\frac{w_{it}}{w_{it}(\hat{\theta}^{(m)})} + \lambda^{(m)} g^{(m)}(\hat{\theta}^{(m)})' \left[\Sigma^{(m)}(\hat{\theta}^{(m)})\right]^{-1} g^{(m)}(\hat{\theta}^{(m)}) \tag{20}$$

where $g^{(m)}(\theta^{(m)}) = R^{(m)}\theta^{(m)} - r^{(m)}$ represents the linear constraints for the parameter $\theta^{(m)}$. $R^{(m)}$ and $r^{(m)}$ are properly defined for such constraints and $\Sigma^{(m)}(\hat{\theta}^{(m)}) = Var\left(g^{(m)}(\hat{\theta}^{(m)})\right)$.

We add the second term with weight $\lambda^{(m)}$ to take account of the impacts of constraints.

We select a consumer demand model which gives the minimum value of the information measure just defined above. Further we apply the bootstrap method to achieve the robust model selection.

The performance of these two measures will be examined in the next section.

4. MONTE CARLO EXPERIMENTS

The object of the simulation is to see the performance of the measures defined in the previous section. Further we will find the optimal weight $\lambda^{(m)}$ for the measure with unconstrained estimator (20). The models we use in this section are Rotterdam, Linear Expenditure and Almost Ideal Demand System (AI) models.

4.1 Simulation

We use the seasonally adjusted quarterly series of the composition of the final consumption expenditure of households in the domestic market from Annual Reports on National Accounts in Japan. The data from the second quarter of 1970 through to the first quarter of 1998 is available and the first observation is utilized only for calculating the difference, so we have 111 observations for the simulation. Total consumption expenditure is divided into 8 categories: 1) Food, beverages and tobacco, 2) Clothing and footwear, 3) Gross rents, water, fuel and power, 4) Furniture, furnishing and household equipment and operation, 5) Medical care and health expenses, 6) Transport and communication, 7) Recreation, entertainment, education and cultural services, and 8) Other goods and services.

First, we estimate several consumer demand system model described in the previous section. Then we use the estimated constrained parameters for data generating process of each model. Once we have the data for simulation, we will proceed the simulation as follows:

(i) estimate all models and obtain the values of the model selection measures defined in section 3 for all models,

(ii) select the model which attains the minimum value of the measure. **(Result M)**

(iii) adopt Bootstrap method for (i)

(iv) select the model which attains the minimum value of the measure. **(Result B)**

The number of simulation is 500 times. In each simulation, we conduct 100 times of resampling.

The selection ratios with the measure $\hat{I}^{(m)}$ in the Result M and Result B are summarized in Table 1 and 2, respectively.

In these tables, h represents the size of the weight:

$$\lambda^{(m)} = (h \times 2 \cdot \mathrm{rank}(R) \times \#\{\text{observations}\})^{-1}$$

where #{observations} means the number of the observations, and bold faced numbers of h in each table are the optimal weight, which maximizes the rate when the right model are selected. These results shows the optimal weights $\lambda^{(m)}$ are different for the measure (20) and its modified version with bootstrap method.

Table 1: The selection ratio with $\hat{I}^{(m)}$: Result M

h	Selected \ True	RT	LE	AI
5	RT	1.00	0.00	0.00
	LE	0.00	0.02	0.00
	AI	0.00	0.98	1.00
4	RT	1.00	0.00	0.00
	LE	0.01	0.12	0.00
	AI	0.00	0.88	1.00
3	RT	0.92	0.00	0.00
	LE	0.08	0.48	0.00
	AI	0.00	0.52	1.00
2.6	RT	0.81	0.00	0.00
	LE	0.19	0.68	0.02
	AI	0.00	0.32	0.98
2	RT	0.44	0.00	0.00
	LE	0.56	0.92	0.13
	AI	0.00	0.08	0.87
1	RT	0.01	0.00	0.00
	LE	0.99	1.00	0.89
	AI	0.00	0.00	0.11

Table 2: The selection ratio with $\hat{I}^{(m)}$: Result B

h	Selected \ True	RT	LE	AI
5	RT	0.99	0.00	0.00
	LE	0.01	0.47	0.00
	AI	0.00	0.53	1.00
4.1	RT	0.92	0.00	0.00
	LE	0.08	0.87	0.01
	AI	0.00	0.13	0.99
4	RT	0.87	0.00	0.00
	LE	0.13	0.90	0.02
	AI	0.00	0.10	0.98
3	RT	0.25	0.00	0.00
	LE	0.75	0.99	0.12
	AI	0.00	0.01	0.88
2	RT	0.00	0.00	0.00
	LE	1.00	1.00	0.99
	AI	0.00	0.00	0.01
1	RT	0.00	0.00	0.00
	LE	1.00	1.00	1.00
	AI	0.00	0.00	0.00

The selection ratios with the measure $\tilde{I}^{(m)}$ in the Result M and Result B are very similar, so we summarize the Result B in Table 3.

Table 3: the selection ratio with $\tilde{I}^{(m)}$

Selected \ True	RT	LE	AI
RT	0.99	0.00	0.00
LE	0.00	0.00	0.00
AIDS	0.01	1.00	1.00

This result shows that the measure $\tilde{I}^{(m)}$ misleads when the true model is the linear expenditure model. Judging from the simulation, we recommend to utilize the modified version of the measure $\hat{I}^{(m)}$ with bootstrap method.

4.2 Empirical Example

We apply the proposed model selection measure to the data used in previous sub-section. Without modification, the measures $\tilde{I}^{(m)}$ and $\hat{I}^{(m)}$ with h=2.6 select the RT model. But, the modified version of these two measures do not select the same model. Table 4 shows the selection ratio in 500 times resampling.

Table 4: the selection ratio

criteria	RT	LE	AI
$\tilde{I}^{(m)}$	1.00	0.00	0.00
$\hat{I}^{(m)}$ with h=4.1	0.00	1.00	0.00

Because $\tilde{I}^{(m)}$ does not perform well and the modified version of the measure $\hat{I}^{(m)}$ selects the true model more frequently than the measure without bootstraping, we conclude that the linear expenditure model should be selected .

5. CONCLUSIONS

In this paper, we propose the model selection measure for the consumer demand systems. This measure is based on the minimum entropy using the estimated budget shares, not based on likelihood. We find the proposed measure works well when the weight $\lambda^{(m)}$ is properly selected. Specially, the modified measure (20) by the bootstrapping performs reasonably well. Utilizing the optimal weight for the constraints and bootstrap method, we can select the true model more frequently.

REFERENCES

Barten, A.P., Consumer Demand Function under Conditions of Almost Additive Preference, *Econometrica*, 32, 1-38, 1964.

Finke, R., and H. Theil, An Extended Version of Minimum Information Estimation of Allocation Models. *Economics Letters*, 15, 229-233, 1984.

Deaton, A. and J. Muellbauer, "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326, 1980.

Stone, J.R.N., Linear Expenditure System and Demand Analysis: An Application to the Pattern of British Demand, *Economic Journal*, 64, 511-527, 1954.