Journal of Agricultural and Applied Economics, Supplement, 35(2003):139–145

The Informational Fit and Maximum Likelihood in a Pooled Cross-Country Demand System with Autocorrelation

Dongling Chen and James L. Seale, Jr.

We fit the Florida Model with an AR(1) error structure to pooled cross-country International Comparison Project (ICP) data of Seale, Walker, and Kim and estimate the model with the minimum information (MI) estimator. Point estimates obtained by MI are similar in value to those obtained by Seale, Walker, and Kim with maximum likelihood (ML). Two similar simulations but with different sample sizes are conducted to compare the relative efficiencies of MI and ML with known and unknown (MLU) covariances. In the larger sample, the MLU is more efficient in terms of root-mean-squared errors (RMSEs) than the MI. Noteworthy, in the small sample, the MI is more efficient in terms of RMSEs than MLU, even though MLU explicitly accounts for AR(1), whereas the MI does not. These results correspond to earlier findings of Theil for time-series and cross-sectional data.

Key Words: autocorrelation, cross-country demand, maximum likelihood, minimum information, pooled data

Theil published four papers in 1984 (i.e., Fink, Floyd, and Theil; Flood, Finke, and Theil; Finke and Theil; Theil, Finke, and Flood) on the informational fit of demand systems. These papers were concerned with the efficiency of the maximum likelihood (ML) estimator in demand systems as well as the informational fit of these systems for individual-country analyses using time-series data.

In 1996, Theil published with Chen a paper, "The Informational Fit of a Cross-Country Demand System," as Chapter 4 of his last book, *Studies in Global Econometrics*. In that chapter, Theil and Chen summarized the results of Theil's previously published work on the informational fit of demand systems and extended that work further by comparing the ML estimator and the minimum information (MI) estimator in a cross-country demand system with cross-sectional data. The chapter also includes discussion of informational inaccuracies and of Stoebel measures as a means to measure the fit of cross-country demand systems and to identify outliers. In the appendices, Theil and Chen discussed elements of statistical information theory, provided a description of the International Comparison Project data for Phases II, III, and IV, discussed the elements of utility theory, and described the estimation procedure.

In the last section of the chapter, Theil and Chen made several suggestions for further research, one of which was the investigation of the efficiency of the ML estimator by comparison with the MI estimator when cross-country data are pooled over time and when persistent

---

Dongling Chen is General Manager of Beijing BYL SVA Network Co., Ltd., working in Beijing. James L. Seale, Jr., is a professor in the Food and Resource Economics Department, University of Florida.

This project was partially funded by ERS, USDA Cooperative Agreements 43-3AEL-7-80076 and 43-3AEL-2-80029 and NRI Cooperative Research Program 9302858. This is Florida Agricultural Experiment Station Journal Series R-09369.

cross-country differences in consumer preferences are taken into account with an autocorrelation process. In the present article, we do just that. Specifically, we fit the Florida model with an AR(1) error structure to the pooled cross-country data of Seale, Walker, and Kim for an 11-good demand system. We estimate the model with the MI estimator and compare our parameter estimates to those obtained by Seale, Walker, and Kim with ML. The MI estimator is covariance free and does not explicitly account for autocorrelation, whereas the ML procedure does. Next, we perform two simulation experiments to compare results from MI and ML with known and unknown covariances. In the first simulation, the pooled data set is relatively large (96 observations), whereas, in the second, the data set is significantly smaller (18 observations).

Estimating the Florida Model

The Florida Model was originally developed by Theil, Chung, and Seale and named the Working preference independence model. Seale, Walker, and Kim renamed it the Florida model at the suggestion of Henri Theil in the tradition of naming differential demand systems by the place at which they were developed (e.g., Rotterdam model, Theil 1965; Central Bureau of Statistics model, Keller and van Driel; and National Bureau of Research model, Neves). Theil also referred to the model in later writing as the Florida model (e.g., Theil 1997; Theil and Chen).

Let w_{ic} be the budget share of good i in country c, q_c the log of real per capita income of country c, p_{ic} the price of good i in country c, and $\bar{p}_i$ the geometric mean of the prices of i in all countries. The Florida model for a cross-country demand system can be written as

$$(1)\quad w_{ic} = \alpha_i + \beta_i q_c + (\alpha_i + \beta_i q_c) \times \left[\log\frac{p_{ic}}{\bar{p}_i} - \sum_{j=1}^{n} (\alpha_j + \beta_j q_c)\log\frac{p_{jc}}{\bar{p}_j}\right] + \varphi(\alpha_i + \beta_i q_c^*) \times \left[\log\frac{p_{ic}}{\bar{p}_i} - \sum_{j=1}^{n} (\alpha_j + \beta_j q_c^*)\log\frac{p_{jc}}{\bar{p}_j}\right] + \varepsilon_{ic}$$

where φ is the income flexibility; $q_c^* = q_c + 1$; and the error vector $(\varepsilon_{1c}, \varepsilon_{2c}, \ldots)$ has mean zeros and covariance Ω.

Model (1), the Florida model, was originally designed for a demand system with N different countries at a specific time. If there are two or more periods, the error term is assumed to follow an AR(1) scheme

$$(2)\quad \varepsilon_{ct} = \tau\varepsilon_{c,t-1} + v_{ct}$$

where ε_{ct} is an error vector of n commodities in country c and time t; $t - 1$ and t refer to two successive periods, whereas v_{ct} is *iid* $N(0, \Omega)$ of order $(n - 1) \times 1$ and $-1 < \tau < 1$.[1] As stated earlier, the AR(1) process accounts for persistent cross-country differences in consumer preferences.

Using data from the International Comparison Project (ICP) for 1970, 1975, and 1980, Seale, Walker, and Kim fitted the Florida model to an 11-good system and estimated it by ML, independently for 1975 and 1980, and jointly for all three periods.[2] The sample sizes for 1975 and 1980 were 30 and 24, respectively, and 96 for the pooled estimation. The ML estimates for the pooled data are represented in column 2 of Table 1 and are taken from column 4 of table 1 in Seale, Walker, and Kim.

ML has well-known optimal properties for large samples under appropriate conditions, but there are always questions regarding ML and the size of the sample. For example, Theil and Chen, among others, found the sampling variances of the ML parameter estimates to increase as the number of equations increases and as the sample size decreases; in these cases, the asymptotic variances have a strong tendency to underestimate the true variances.

The MI procedure is as follows (Theil and Chen). Assume that suitable data exist for N countries ($c = 1, \ldots, N$). Let θ be the parameter vector of the model and write $\bar{w}_{ic}(\theta)$ for the predicted per capita budget share of good

[1] Because of the adding-up restriction, we estimate the demand system with only $n - 1$ equations (Barten).

[2] See the appendix in Seale, Walker, and Kim for a description of the ML estimation procedure with autocorrelation and unbalanced pooled data.

Table 1. Pooled Estimates of the Florida Model (Asymptotic Standard Errors in Parentheses)

Good or Parameter (1)	ML* (2)	MI† (3)
Income flexibility ϕ	−0.664 (0.026)	−0.589
Coefficient β_i		
Food	−0.139 (0.008)	−0.142
Beverages, tobacco	0.001 (0.004)	−0.001
Clothing, footwear	−0.006 (0.003)	−0.003
Gross rent, fuel	0.019 (0.004)	0.024
Energy	0.015 (0.003)	0.015
House furnishing	0.018 (0.003)	0.022
Medical care	0.019 (0.003)	0.024
Transport, communications	0.023 (0.005)	0.016
Recreation	0.018 (0.003)	0.019
Education	0.002 (0.005)	−0.004
Other	0.030 (0.005)	0.031
Coefficient α_i		
Food	0.171 (0.010)	0.163
Beverages, tobacco	0.057 (0.005)	0.055
Clothing, footwear	0.080 (0.004)	0.081
Gross rent, fuel	0.102 (0.005)	0.107
Energy	0.069 (0.004)	0.068
House furnishing	0.089 (0.004)	0.093
Medical care	0.084 (0.004)	0.089
Transport, communications	0.095 (0.006)	0.089
Recreation	0.066 (0.003)	0.069
Education	0.063 (0.005)	0.059
Other	0.124 (0.006)	0.126
τ	0.711 (0.022)	NA‡

* ML is maximum likelihood. The estimates in this column are from Seale, Walker and Kim, Table 1, column 4, p. 36.
† MI is minimum information.
‡ NA = not applicable.

i in country c when the model uses θ as a parameter vector. Then

$$(3)\qquad I_c(\theta) = \sum_{i=1}^{n} w_{ic}\log\left[\frac{w_{ic}}{\hat{w}_{ic}(\theta)}\right]$$

is the information inaccuracy for country c given θ. Naturally, we are interested in minimizing this inaccuracy, at least on average, over the N countries of the sample. Therefore, the MI estimate of θ is obtained by minimizing

$$(4)\qquad \bar{I}_c(\theta) = \left(\frac{1}{N}\right)\sum_{c=1}^{N} I_c(\theta)$$

with respect to θ.

The Florida model with AR(1) specification is fit to the pooled cross-country data of Seale, Walter, and Kim and parameter estimates are obtained with MI. These are presented in column 3 of Table 1. Some differences between the ML estimates (column 2) and the MI estimates (column 3) of Table 1 exist. In particular, the estimate of the income flexibility is −0.66 for ML and −0.59 for MI. Moreover, the ML procedure produces asymptotic standard errors for its estimates, whereas MI does not.[3] Because the MI procedure is distribution free, the AR(1) coefficient, τ of Equation (2), is not involved in the MI procedure.

[3] As shown below, however, root-mean-squared errors can be calculated for the MI parameter estimates.

A First Simulation Experiment

To compare the two different approaches, we designed a simulation experiment in which Ω is known, so that three estimation procedures are available: ML with known Ω (MLK), ML with unknown Ω (MLU), and MI. The simulation design uses true values of the coefficients (α_is, β_is, φ, and τ) and of the covariance matrix Ω. The simulation proceeds as follows: (1) generate a pseudo-normal random vector with zero mean and three kinds of covariances,

$$\Omega_1 = \Omega \quad \text{if a country participates in only one phase,}$$

$$\Omega_2 = \begin{bmatrix} \Omega & \tau\Omega \\ \tau\Omega & \Omega \end{bmatrix} \quad \text{if a country participates in two phases, and}$$

$$\Omega_3 = \begin{bmatrix} \Omega & \tau\Omega & \tau^2\Omega \\ \tau\Omega & \Omega & \tau\Omega \\ \tau^2\Omega & \tau\Omega & \Omega \end{bmatrix} \quad \text{if a country participates in three phases;}$$

(2) use the true values of the coefficients, the simulated error vector and model Equation (1) with Equation (2) to form the simulated value of the dependent variables; and (3) estimate Equation (1) with Equation (2) using the observed independent variables and the simulated dependent variables with three methods: MLK, MLU, and MI. We repeated this procedure 1,000 times.

The arithmetic means of these estimates (not shown here) are all close to the true values; bias does not appear to be a problem. Table 2 presents the root-mean-squared errors (RMSEs) and the root-mean-asymptotic-squared errors (RMASEs) of the estimates for 1,000 trials. As can be seen, columns 2 and 3 show the true values and the true asymptotic standard errors of the estimates; columns 4, 6, and 8 give the RMSEs around the true values for MLU, MLK, and MI, respectively; and columns 5 and 7 present the RMASEs of the estimates for MLU and MLK, respectively. Comparing columns 4 and 5, all RMASEs for MLU are smaller than the corresponding MLU RMSEs and the true asymptotic standard errors (except in the case of τ), which indicates the estimated asymptotic standard errors understate the true asymptotic standard errors when the covariance matrix is unknown and must be estimated. The RMASEs and RMSEs in columns 6 and 7 for MLK are close to each other and are close to the true asymptotic standard errors (column 3). A comparison of columns 4 and 6 indicates that MLK performs better than MLU, because the pairwise RMSEs for MLK are smaller than those for MLU. These results agree with many previous findings (e.g., Theil and Chen; Theil, Chung, and Seale). However, the RMSEs in column 8 for MI are substantially greater than the corresponding ones in columns 4 and 6 for MLU and MLK, respectively, which implies that MI performs the worst among the three estimators in this particular sample.

The relatively poor performance of the MI estimator versus the MLU and especially the MLK estimators could be due to both MLU and MLK explicitly taking into account the imposed AR(1) structure of the error term whereas the MI estimator is distribution free and does not explicitly or implicitly take the error structure into account. The other reason could be due to the relatively large sample size of the pooled data (96 observations). Theil and Chen (1995) show that MI performs better than MLU but worse than MLK for sample sizes $\leq$30. They also show that the superior performance of MI relative to MLU diminishes as the sample size increases.

A Second Simulation Experiment

In the previous section, we found that both MLU and MLK perform better than MI for the pooled data set when the Florida model is specified with an AR(1) structure. In this section, we develop a simulation experiment to test whether the above results are due to the autocorrelation specification or the relatively large sample size (96 observations). Specifically, we repeat the above simulation but with a much smaller sample size and retain the AR(1) specification for MLU and MLK.

Table 2. A Simulation Experiment for 11 Goods and Pooled Data with 96 Observations

Coefficient or Good (1)	True Value (2)	Asymptotic Standard Error (3)	MLU* RMSE (4)	MLU* RMASE‖ (5)	MLK† RMSE	MLK† RMASE	MI‡ RMSE (8)
Income flexibility ϕ	−6,639	261	284	237	265	270	386
Coefficient β_i							
Food	1,387	78	80	74	78	78	87
Beverages, tobacco	8	37	37	35	36	37	41
Clothing, footwear	60	30	31	29	30	30	35
Gross rent, fuel	193	36	37	34	37	37	45
Energy	152	28	29	26	28	28	33
House furnishing	181	33	35	31	34	33	39
Medical care	192	32	33	30	32	32	40
Transport, communications	232	45	46	43	45	45	52
Recreation	176	25	24	24	24	25	27
Education	19	53	56	49	54	54	67
Other	295	46	47	44	45	46	54
Coefficient α_i							
Food	1,705	101	105	97	104	99	110
Beverages, tobacco	568	55	56	52	55	54	59
Clothing, footwear	799	41	42	39	42	40	45
Gross rent, fuel	1,022	53	52	50	52	52	56
Energy	691	40	40	38	40	39	44
House furnishing	894	43	45	42	45	43	49
Medical care	844	41	41	39	41	41	46
Transport, communications	946	57	58	55	57	57	62
Recreation	657	34	33	32	33	33	35
Education	634	50	52	47	51	51	60
Other	1,236	61	65	59	64	61	69
τ	7,109	22	32	24	61	26	NA¶

Note: All entries to be divided by 10,000.
* MLU = Maximum likelihood with unknown Ω.
† MLK = Maximum likelihood with known Ω.
‡ MI = minimum information.
§ RMSE = root-mean-square error.
‖ RMASE = root-mean-asymptotic-square error.
¶ NA = not applicable.

We now use only the first 6 of 13 countries that have three phases and the sample size for this simulation is now equal to 3 × 6 = 18 observations, which is significantly smaller than the original size, 96. The simulation procedure is the same as the previous one except that only Ω_3 is used for this simulation.

The simulation results are given in Table 3. Comparing Table 3 with Table 2, we can see that all the RMSEs have significantly increased, but those of the MLU increase the most. The MLU-RMSE of the income flexibility in Table 3 is almost four times larger than the corresponding one in Table 2, whereas the MLU-RMSEs of all other coefficients are more than double the corresponding ones in Table 2. The RMSEs for MLK and MI in columns 4 and 8 of Table 3 have also increased with the decreased sample size, but their RMSEs of the coefficients are all smaller than the corresponding MLU ones in Table 3. This is as expected and corresponds to earlier results of Theil and his coauthors.

Of particular interest to us is the relative

Table 3. A simulation Experiment for 11 Goods and Pooled Data with 18 Observations

Coefficient or Good (1)	True Value (2)	Asymptotic Standard Error (3)	MLU* RMSE§ (4)	MLU* RMASE‖ (5)	MLK† RMSE (6)	MLK† RMASE (7)	MI‡ RMSE (8)
Income flexibility ϕ	−6,639	261	1,072	825	605	604	908
Coefficient β_i							
Food	1,387	78	207	93	143	140	156
Beverages, tobacco	8	37	84	40	62	618	66
Clothing, footwear	60	30	72	33	51	50	55
Gross rent, fuel	193	36	91	41	63	65	78
Energy	152	28	74	33	52	52	56
House furnishing	181	33	79	36	53	54	57
Medical care	192	32	94	42	66	63	71
Transport, communications	232	45	113	53	74	79	78
Recreation	176	25	63	30	44	45	47
Education	19	53	200	67	121	126	165
Other	295	46	111	53	73	78	80
Coefficient α_i							
Food	1,705	101	226	140	190	182	197
Beverages, tobacco	568	55	111	76	101	99	103
Clothing, footwear	799	41	88	57	77	73	79
Gross rent, fuel	1,022	53	117	75	101	97	109
Energy	691	40	88	56	79	72	82
House furnishing	894	43	96	60	80	77	82
Medical care	844	41	96	60	84	77	87
Transport, communications	946	57	133	79	106	103	110
Recreation	657	34	77	47	67	61	69
Education	634	50	156	67	104	106	135
Other	1,236	61	134	85	115	110	119
τ	7,109	22	193	75	107	71	NA¶

Note: All entries to be divided by 10,000.
* MLU = Maximum likelihood with unknown Ω.
† MLK = Maximum likelihood with known Ω.
‡ MI = minimum information.
§ RMSE = root-mean-square error.
‖ RMASE = root-mean asymptotic-square error.
¶NA = not applicable.

performance of MLU, MLK, and MI in this smaller sample size. The results imply that MLU with an autoregressive process performs poorly in small samples. Pairwise, the MLU-RMSEs are all larger than the corresponding ones of MLK and MI, whereas those of MLK are pairwise the smallest. Accordingly, MI outperforms the MLU and the performance of MLK is again best; however, knowledge of the covariance matrix is exceptional in empirical work. That the MI estimator, in a small sample, outperforms MLU in a pooled cross-country demand system with autocorrelation is certainly noteworthy, and this is the case, even though the MLU explicitly accounts for autocorrelation in the likelihood function whereas the MI does not.

Conclusions

In the present article, we fit the Florida model with an AR(1) error structure to the pooled cross-country data of Seale, Walker, and Kim and estimate the model with MI. Our point estimates obtained by MI are similar in value to the ML values of Seale, Walker, and Kim. Two similar simulations, but with different sample sizes, are conducted to compare the relative efficiencies of the MLU, MLK, and MI. As expected, the MLU is more efficient in terms of RMSEs than the MI in the larger sample. Noteworthy, the MI is more efficient in terms of RMSEs than the MLU in the small sample. This is the case even though the MLU explicitly accounts for the AR(1) error structure whereas the MI, a distribution-free estimator, does not. These results for pooled cross-country data correspond to earlier findings of Theil for time-series and cross-sectional data.

References

Barten, A.P. "Maximum Likelihood Estimation of a Complete System of Demand Equations." *European Economics Review* 1(1969):7–73.

Finke, R., L.R. Flood, and H. Theil. "Minimum Information Estimation of Allocation Models of Different Sizes." *Statistics and Probability Letters* 2(1984):279–83.

Finke, R., and H. Theil. "An Extended Version of Minimum Information Estimation of Allocation Models." *Economics Letters* 15(1984):229–33.

Flood, L.R., R. Finke, and H. Theil. "Maximum Likelihood and Minimum Information Estimation of Allocation Models with Fat-Tailed Error Distribution." *Economics Letters* 16(1984): 213–8.

Keller, W.J., and J. van Driel. "Differential Consumer Demand Systems." *European Economic Review* 27(1985):375–90.

Neves, P. "Analysis of Consumer Demand in Portugal, 1958–1981." Memoire de Maitrise en Sciences Economiques, Universite Catholique de Louvain, Louvain-la-Neuve, 1987.

Seale, J.L., Jr., W.E. Walker, and I.-M. Kim. "The Demand for Energy: Cross-Country Evidence Using the Florida Model." *Energy Economics* 13(1991):33–9.

Theil, H. "The Information Approach to Demand Analysis." *Econometrica* 33(1965):67–87.

Theil, H. "The Florida Model for Demand Analysis with Cross-Country Data." *Price Policies and Economic Growth.* A. Jorge and J. Salazar-Carillo, eds. Westport, CN: Praeger, 1997.

Theil, H., and D. Chen. "The Informational Fitting of a Cross-Country Demand Model." *Studies in Global Econometrics.* Henri Theil. Westport, CN: Praeger, 1996.

Theil, H., C.-F. Chung, and J.L. Seale, Jr. *International Evidence on Consumption Patterns.* Greenwich, CN: JAI Press Inc., 1989.

Theil, H., R. Finke, and L.R. Flood. "Minimum Information Estimation of Allocation Models." *Economics Letters* 15(1984):251–6.