Review of Income and Wealth
Series 49, Number 3, September 2003

RECONCILING HOUSEHOLD SURVEYS AND NATIONAL ACCOUNTS DATA USING A CROSS ENTROPY ESTIMATION METHOD

BY ANNE-SOPHIE ROBILLIARD*

IRD-DIAL, Paris

AND

SHERMAN ROBINSON

IFPRI, Washington DC

This paper presents an approach to reconciling household surveys and national accounts data. The problem is how to use the information provided by the national accounts data to re-estimate the household weights used in the survey so that the survey results are consistent with the aggregate data. The estimation approach uses an estimation criterion based on an entropy measure of information. The survey household weights are treated as a prior. New weights are estimated that are close to the prior and that are also consistent with the additional information. This approach is implemented to reconcile household survey data and macro data for Madagascar. The results indicate that the approach is powerful and flexible, supporting the efficient use of information from a variety of sources to reconcile data at different levels of aggregation in a consistent framework.

1. INTRODUCTION

Reconciling household survey data and national accounts data is a well-known problem. Computing macro aggregates from household survey data by multiplying household production, income, consumption, and/or savings by the household sample weights and summing virtually never matches published national accounts data, even though the sample weights are designed to represent the national population. Many reasons are offered to explain this mismatch. On the household survey side, there may be sampling errors due to inadequate survey design and/or measurement errors because it is difficult to get accurate responses from households concerning economic variables. On the national accounts side, while supply-side information on output and income for some production sectors is based on high-quality survey or census data for agriculture and industry, information for subsistence farmers and informal producers is harder to obtain and usually of lower quality.

For many purposes, it is important to be able to reconcile household surveys and national accounts data. Policy implications drawn from analysis of household

Note: The authors wish to thank INSTAT and MADIO for making the data available, and participants at seminars at Yale University and IFPRI as well as two anonymous referees for helpful comments and suggestions.

*Correspondence to: Anne-Sophie Robilliard, Chargée de Recherche, DIAL-IRD, 4 rue d'Enghien, 75010 Paris, France (robilliard@dial.prd.fr).

surveys may well give misleading implications about aggregate costs of a given policy initiative if the survey results do not accurately "blow up" to national aggregates. Similarly, it is often desirable to disaggregate the national data to incorporate greater sectoral, regional, or household detail (van Tongeren, 1986). The goal is to use household survey data to provide the basis for such disaggregation, usually in the framework of a social accounting matrix (SAM), which provides a consistent accounting system for reconciling national, regional, and household accounts. Finally, there is a strand of work using household survey data to provide the foundation for microsimulation models that specify the behavior of each household and simulate their interactions across markets. If such models are to provide an adequate framework for policy analysis, it would be ". . . helpful if the national accounts aggregates are consistent with the microsimulations" (Pyatt, 1991).

In this paper, we present an approach to reconciling household surveys and national accounts data that starts from the assumption that the macro data represent control totals to which the household data must be reconciled. We will also assume that the economic data gathered in the survey are accurate, or have been adjusted to be accurate. The first assumption will then be relaxed, and an "errors in aggregates" version of the problem will be presented as well.[1] Given these assumptions, the problem is how to use the additional information provided by the national accounts data to re-estimate the household weights used in the survey so that the survey results are consistent with the aggregate data, while simultaneously estimating the errors in the aggregates. The approach we take represents an efficient "information processing rule" that uses an estimation criterion based on an entropy measure of information. The results indicate that the approach is powerful and flexible, supporting the efficient use of information from a variety of sources to reconcile data at different levels of aggregation in a consistent framework.

Section 2 presents the background and a mathematical description of the estimation problem, while Section 3 presents an application to the case of Madagascar.

2. Information Theory and Parameter Estimation

The starting point for the estimation approach is information theory as developed by Shannon (1948) and applied to problems of estimation and statistical inference by Jaynes (1957). The philosophy underlying this approach is to use *all*, and *only*, the information available for the estimation problem at hand. Our goal is to estimate a set of household survey weights consistent with extraneous supply-side information in the form of national accounts data. Two types of information are available for our purpose. First, sample design is a major effort in any household survey and the estimated household weights resulting from this effort embody a lot of demographic information. These weights should provide a starting point for any estimation procedure. In our approach, we use these weights as a "prior"

[1]See Golan, Judge, and Miller (1996) and Robinson, Cattaneo, and El-Said (2001) for "errors in variables" and "errors in equations" applications of the cross entropy estimation approach.

and estimate new coefficients that are "close" to the prior but are consistent with other information. The second type of information comes from two sources: the results of the household survey and independently generated data from other sources such as the national accounts and/or other surveys. This second type of information can be expressed in the form of known weighted averages or "moments" of the distribution of observed variables across the households in the sample.

The estimation problem can be restated as follows. Estimate a set of sampling probabilities (household survey weights) that are close to a known prior and that satisfy various known moment constraints. Consider a sample survey of K households with prior survey probabilities $\bar{p}_k$ which results in a vector x_k of observed characteristics for each household such as household size, total household income, income by source, consumption, and so forth. In addition, from other sources, we have information about aggregations or weighted averages of some of the household information. The estimation procedure is to minimize the Kullback-Leibler cross entropy measure of the distance between the new estimated probabilities and the prior. Following the notation of Golan, Judge, and Miller (1996), the estimation procedure is:

$$\text{(1)} \qquad \text{Min} \sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{\bar{p}_k}\right)$$

subject to moment consistency constraints

$$\text{(2)} \qquad \sum_{k=1}^{K} p_k f_t(x_k) = y_t \quad t \in [1, \ldots, T]$$

and the adding-up normalization constraint

$$\text{(3)} \qquad \sum_{k=1}^{K} p_k = 1$$

where $\{y_1, y_2, \ldots, y_T\}$ is an observed set of data (e.g. averages or aggregates) that is required to be consistent with the distribution of probabilities or sample frequencies (weights) $\{p_1, p_2, \ldots, p_K\}$. The function f_t represents a general aggregator of within-household variables. In our case, the function simply picks out a particular variable and we could have replaced it with the observations $x_{t,k}$. K is usually very large, in the thousands, while T is small, representing a few macroeconomic and demographic adding-up constraints. In terms of classical statistical parameter estimation, the problem is undetermined or "ill posed." There are not enough degrees of freedom to support estimation. The cross entropy approach uses all available information, including prior parameter estimates, and supports estimation even in a "data sparse" environment.

The use of the cross entropy measure in the estimation criterion has been justified on the basis of axiomatic arguments concerning its desirability both as a measure of "information" and as a criterion for inference.[2] There are close links between the minimum cross entropy criterion and maximum likelihood estimators,

[2]See Kapur and Kesavan (1992) and Golan, Judge, and Miller (1996).

but the cross entropy criterion requires fewer statistical assumptions in that its application does not require specification of an explicit likelihood function.[3] In our case, this sparseness in assumptions is desirable since we have no knowledge about the form of any underlying probability distributions.

The probability weights are estimated by minimizing the Lagrangian:

$$L=\sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{\bar{p}_k}\right)+\sum_{t=1}^{T}\lambda_t\left(y_t-\sum_{k=1}^{K}p_k f_t(x_k)\right)+\mu\left(1-\sum_{k=1}^{K}p_k\right) \tag{4}$$

The first-order conditions are:

$$\frac{\partial L}{\partial p_k}=\ln p_k-\ln\bar{p}_k+1-\sum_{t=1}^{T}\lambda_t f_t(x_k)-\mu=0,\quad k\in[1,\ldots,K] \tag{5}$$

$$\frac{\partial L}{\partial \lambda_t}=y_t-\sum_{k=1}^{K}p_k f_t(x_k)=0,\quad t\in[1,\ldots,T] \tag{6}$$

$$\frac{\partial L}{\partial \mu}=1-\sum_{k=1}^{K}p_k=0 \tag{7}$$

The solution can be written as:

$$\tilde{p}_k=\frac{\bar{p}_k}{\Omega(\tilde{\lambda}_1,\tilde{\lambda}_2,\ldots,\tilde{\lambda}_T)}\exp\left[\sum_{t=1}^{T}\tilde{\lambda}_t f_t(x_k)\right],\quad k\in[1,\ldots,K] \tag{8}$$

where

$$\Omega(\tilde{\lambda})=\sum_{k=1}^{K}\bar{p}_k\exp\left[\sum_{t=1}^{T}\tilde{\lambda}_t f_t(x_k)\right] \tag{9}$$

is defined as the "partition function" and ensures that the estimated probabilities sum to one.

The solution equation (8) shows how estimated weights depend on prior weights and constraints. If none of the constraints were binding, then all the lambdas would be zero, and the estimated weights would be equal to their prior (since the sum of the p_k is equal to one). In this situation, the moment constraints add no information to the estimation problem. If constraints are binding, then the estimated weights depend on the prior, the value of the lambdas, and the value of the variables $f_t(x_k)$ associated with the constraints.

We now generalize our approach to the case where macro aggregates are not exact but are measured with error. We start by assuming that we have some knowledge about the standard error (perhaps due to measurement error), which we treat as a Bayesian prior, not a maintained hypothesis. The estimated error is specified as a weighted sum of elements in an error support set:

$$e_t=\sum_{t} w_{t,l}\bar{v}_{t,l} \tag{10}$$

[3]See Golan (1998) and Zellner (1990).

where:

e_t = error value
$w_{t,l}$ = error weights estimated in the CE procedure
$\bar{v}_{t,l}$ = error support set

The set l defines the dimension of the support set for the error distribution and the number of weights that must be estimated for each error. The prior variance of these errors is given by:

$$\sigma^2 = \sum_l \bar{w}_{i,l} \cdot \bar{v}_{i,l}^2 \tag{11}$$

where $\bar{w}_{i,l}$ = prior weights on the error support set.

Starting with a prior σ, Golan, Judge, and Miller (1996) suggest picking the $\bar{v}_s$ to define a domain for the support set of ±3 standard errors. In this case, the prior on the weights, $\bar{w}$, are then calculated to yield a consistent prior on the standard error, σ.

With errors in aggregates, the constraints of the problem are

$$\sum_k p_k f_t(x_k) = y_t + \sum_l w_{t,l}\bar{v}_{t,l} \quad t \in [1,\dots,T] \text{ and } l \in [1,\dots,L] \tag{12}$$

and additional adding-up constraints on the error weights

$$\sum_l w_{t,l} = 1 \quad t \in [1,\dots,T] \tag{13}$$

The maximand will now include a new term in the error weights:

$$\sum_{t,l} w_{t,l} \ln\left(\frac{w_{t,l}}{\bar{w}_{t,l}}\right) \tag{14}$$

First order conditions need to be rewritten to take into account these changes. Values of the support set $\bar{v}_{t,l}$ also need to be specified. This identification depends on the domain of the support set and the assumed prior distribution of errors. Assuming a prior distribution with zero mean and a standard error equal to σ, we used a support set with five terms equal to (-3σ, $-\sigma$, 0, σ, 3σ). Assuming normality of the prior distribution, the prior values of the weights can be computed given only knowledge of the prior mean and standard error.[4]

The estimation problem has no closed-form solution, so we must solve it numerically. Unlike the standard linear regression model, where the solution requires only information about various moments of the data (variance and covariance matrices), the estimation problem here uses all the data. The solution can be seen in a Bayesian perspective, although there is no explicit likelihood

[4]We start with a known mean and variance, and also that the value of kurtosis for the normal distribution is a function of the variance. This approach applied to Social Accounting Matrices is described in Robinson, Cattaneo and El-Said (2001). It is similar to a Bayesian approach to estimating National Accounts developed by Magnus, van Tongeren and de Vos (2000). In both approaches, the posterior includes revised estimates of the moments of the error distributions (mean, variance, and, in our case, kurtosis).

function. The estimation procedure "adjusts" the prior probabilities using the new information to generate posterior estimates. Zellner (1988) calls this procedure an efficient "information processing rule" in that it uses all the information available but does not introduce any assumptions about information we do not have. The use of a different objective function will violate this criterion, implicitly assuming additional unwarranted information in the estimation. Traditional grossing-up methods such as imposing proportional changes implicitly assume an enormous amount of prior information.[5]

3. Reconciling LSMS Survey Data and Macro Data for Madagascar

To illustrate the cross entropy method, we apply it to reconcile household and macro data for Madagascar. The household data come from a "Living Standards Measurement Survey" (LSMS) for Madagascar called EPM 93 (Enquête Permanente auprès des Ménages). The macro aggregates come from a social accounting matrix (SAM). The resulting reweighted sample is to be used as the starting point of a microsimulation model (Cogneau and Robilliard, 1999).

Survey Data

The EPM survey for the year 1993 is a LSMS survey on 4,508 households which was implemented for the Malagasy state by the INSTAT (Institut National de la Statistique) under the supervision of the PNUD and the World Bank (INSTAT, 1993). It includes a large number of variables. We focus on data concerning demographic composition of the family, employment, time use, agricultural factors of production, activities, expenditures, informal income sources, transfers, and other types of income.

Income sources are aggregated into four types: agricultural, informal, formal, and others. Agricultural income includes income from production of crops (both sold and/or home-consumed), income from livestock (computed as a fixed share of total livestock value plus income derived from sold and/or home-consumed animal products) and income from sharecropping. Informal income is derived from both informal wage labor and self-employment in non-agricultural activities. Formal income is derived from formal wage labor and formal capital income for stockholders. Other sources of income include transfers, either from the government or from other households. For households owning their house, rents are imputed on the basis of a predicted rent derived from a regression of rents paid by tenant households over housing characteristics. Some of these characteristics are also used to determine whether imputed rents are to be considered formal or informal income.

Adjusting Income Data

The approach we take is first to use traditional *ad hoc* techniques to adjust the household data for under reporting of some types of income and then, second, to reweight the households using the cross entropy technique to match known

[5]Grossing-up rules could be imposed as constraints in the cross entropy approach. These techniques can therefore be seen as special cases of the general information theory approach.

aggregates (measured with error). In principle, some of these *ad hoc* adjustments could have been incorporated directly in the cross entropy estimation problem by adding more measurement error equations. The resulting cross entropy minimand would include a number of additional terms. However, since we wanted to focus on the reweighting issue, we decided to use this two-stage procedure.

In our sample, 50 percent of all households report an income lower than their expenditures. This discrepancy can be explained by over reporting of expenditures, under reporting of income, and/or transitory low income due to some temporary shock such as loss of employment or a crop failure. We assume that expenditure data are accurate and focus on the income data. First, adjustments are made for specific types of income. Sharecropping income is assumed to be under reported by all landlords and is inflated to meet the aggregated value of payments made by sharecroppers. For stockholders, formal capital income is adjusted to reproduce the structure of formal income derived form the National Accounts, given labor income derived from formal wage labor. Since these adjustments appear not to be sufficient to fill the gap between income and expenditures, the permanent income approach has been used for those households whose incomes are less than expenditures. The assumption made is that the gap is due to transitory low income and that consumption smoothing (through dissaving and/or borrowing) will allow these households to meet their expenditures. All sources of income are adjusted accordingly. Since the data were collected in 1993, and we need to reconcile them with aggregated income data for 1995, an inflation rate of 207 percent corresponding to the rise in the Consumer Price Index between 1993 and 1995, was applied uniformly to all incomes and expenditures, although one can arguably point out that inflation rates differ between regions. Finally, households with no expenditures or no income, or declaring incomes "too high," are discarded and the final sample has 4,458 households.

The SAM for 1995 is a social accounting matrix with 28 production sectors constructed to support computable general equilibrium (CGE) modeling (Razafindrakoto and Roubaud, 1997). For our purpose, we use an aggregated version of the SAM with only three production sectors corresponding to the three sources of income used to summarize household income information (agricultural, informal, and formal). The main information used is the structure of value added actually paid to households. This includes labor and capital value-added. For the agricultural and the informal production sectors, the amount of value added paid to labor, capital, and land appearing in the SAM corresponds to what households actually earn. Concerning the formal production sector, all labor value added goes to households but non-distributed profits are not taken into account when matching micro and macro data as they are not counted as part of income in the household survey.

The comparison of the information derived from the two sources reveals two main differences (Table 1). First, the weighted sum of household incomes falls short by 15.2 percent compared to the SAM figure. Second, the share of informal income in total income appears overestimated in the household survey compared to the SAM, at the expense of the share of informal income, both from labor and capital.

Extraneous information on population growth as well as its distribution between rural and urban areas has been used to recover demographic figures con-

TABLE 1

COMPARING INFORMATION DERIVED FROM MICRO AND MACRO SOURCES

	EPM 93(1)	SAM 95(2)
Total household income (millions of 95 Franc Malagasy)	9,348	11,400
Mean per capita income (thousands of 95 Franc Malagasy)	751	866
Shares of total income (percent)		
Agricultural income	34.7	36.3
Informal income	30.5	17.4
Formal labor income	12.3	19.4
Formal capital income	13.1	22.5
Exogenous income	9.4	4.4

Notes:
(1) After all adjustments described in text.
(2) See Razafindrakoto and Roubaud (1997).

sistent with the year 1995. It is known from other sources that the annual rate of population growth is 2.9 percent. We assumed that this growth did not change the mean size of households, so that the number of households grows at the same rate as population. Concerning population distribution between rural and urban areas, we assumed that the share of population living in rural areas is 75 percent.

Estimating Household Weights

The estimation procedure is implemented with the GAMS software (Brooke, Kendrick, and Meeraus, 1998) and amounts to solving a non-linear program (NLP) with the cross entropy maximand including the terms from equations (1) and (14) subject to adding-up and moment constraints.[6] The input information is, on the micro side, household characteristics such as size, mean age, gender composition, area (urban/rural), total income, and shares of agricultural income, informal income, formal labor income, formal capital income, and share of other sources of income. The survey weights used as priors are also included in the micro database. On the macro side, information is scarce given the stylized structure of the SAM and consists of the structure of income derived from the SAM 95, population size, and number of households in 1995 (derived from 1993 given population growth). Macro and demographic information are introduced as a set of moment constraints. Different sets of constraints are used (see below). The resulting minimization problem has 4,500 variables and seven to nine moment constraint equations. Recent advances in NLP solvers, which incorporate second derivatives, have made the solution of such large problems routine.[7]

Results

Different strategies have been followed in order to reconcile aggregated household income derived from the EPM 93 and income derived from the SAM 95. We

[6]The problem could probably be solved using any software that can handle NLP problems; however GAMS is probably the best available.

[7]The new solvers are called CONOPT3 and PATHNLP and are described on the GAMS website (http://www.gams.com/). The GAMS code for this problem is available from the authors.

TABLE 2

NEW WEIGHTS DISTRIBUTION

	Prior	FOM	SOM	EIA
Mean weight	557.0	593.7	593.7	593.7
Standard deviation	365.6	402.0	437.9	389.7
Minimum weight	114	0	0	0
Maximum weight	1,909	3,598	3,970	3,365
Number of zero weights	0	129	149	75

Note: FOM = First Order Moments; SOM = Second Order Moments; EIA = Errors in Aggregates.

start by assuming that all household incomes are underestimated uniformly which is reasonable given the high inflation rate in that period, and adjust all households incomes by 15.2 percent prior to running the procedure. The estimation procedure then "works" to estimate weights consistent with the income structure derived from the SAM 95.

While the first two simulations assume perfect information on aggregate values, the third takes into account "errors in aggregates" (EIA). Two sets of constraints are used for household incomes. The first set contains only first order moments constraints for both rural and urban area mean per capita income (FOM), while the second includes second order moments as well for both areas (SOM).[8] Results show that inclusion of the second order moments leads to results that are more satisfactory in terms of income distribution.

In terms of the distribution of weights, Table 2 shows that the results do not appear dramatically different from the prior. The mean weight increases by 6.6 percent as a result of population growth (the underlying assumption being that household size remains constant), while the standard deviation from the mean increases by 6.4 to 13.0 percent. The more significant result is that some weights drop to zero, essentially dropping those households from the sample.[9] As a result, the new samples are smaller. Far fewer households are dropped in the version that incorporates errors in measurement of macro aggregates. This result strongly suggests that reconciliation and estimation methods should simultaneously work at both aggregate and household levels.

Concerning the macro and demographic constraints, results in Table 3 show that the estimation procedure achieves consistency with macro and demographic aggregates. Other demographic indicators are presented to control whether the new samples have been distorted. The estimation procedure appears to leave both the gender balance and the average age unchanged. This demographic information could have been used as constraints had the results changed these balances too much but since it is not required, we preferred to keep the problem as small as possible.

Finally, we are concerned about the impact of the procedure on measured income distribution. Results in terms of percentile ratios (Table 4) show that

[8]Since the survey design is characterized by sample stratification, moment constraints on income are applied for each stratum (urban and rural areas) independently and not over the whole sample.

[9]Dropped households are characterized by high shares of informal and exogenous income, as well as high total incomes compared to the rest the sample.

TABLE 3

SELECTED AGGREGATE RESULTS

	Prior	FOM	SOM	EIA
Total number of households ('000)*	2,649	2,649	2,649	2,649
Total population ('000)*	13,059	13,058	13,058	13,061
Total income (millions of 95 FMG)*	11,315	11,315	11,315	11,567
Mean per capita income ('000 of 95 FMG)*	866	866	866	886
Share rural population (%)*	82.8	75.2	74.7	74.8
Share males (%)	49.5	49.8	50.0	49.8
Mean age (years)	21.5	21.4	21.5	21.6

Notes: *Used as constraints in the program.
FOM = First Order Moments; SOM = Second Order Moments; EIA = Errors in Aggregates.

TABLE 4

PERCENTILE RATIOS FOR DISTRIBUTION OF INCOME PER CAPITA

	Prior	FOM	SOM	EIA
P90/P10	9.575	8.792	8.949	9.032
P90/P50	3.264	3.033	3.077	3.080
P75/P25	3.033	2.967	2.961	2.987
P75/P50	1.751	1.730	1.706	1.716

Note: FOM = First Order Moments; SOM = Second Order Moments; EIA = Errors in Aggregates.

TABLE 5

INEQUALITY MEASURES

	Prior	FOM	SOM	EIA
Theil Index	58.8	69.7	50.7	55.3
Between	7.3	8.9	8.5	8.0
Within	51.5	60.8	42.2	47.3
Theil Index for urban area	62.8	80.8	47.5	55.5
Theil Index for rural area	46.0	44.7	37.9	41.0
Gini Index	52.7	54.4	50.4	51.7

Note: FOM = First Order Moments; SOM = Second Order Moments; EIA = Errors in Aggregates.

relative income distribution does not change dramatically in the reweighted samples. However, Gini and Theil indexes (Table 5) show more sensitivity to the reweighing procedures. In particular, imposing only first order moment constraints (FOM) leads to an strong increase in urban inequality. This can be explained by the fact that the set of constraints leads to an increase in the weights of households that derive a high share of their income from formal production sectors, either through formal wage labor or formal capital stocks. These households are typically found at the top of the distribution. Imposing second moment constraints prevent this increase in inequality. In terms of income distribution, introducing errors in aggregated "EIA" appears more satisfactory since the resulting income distribution is close to the prior.

4. Conclusion

The cross entropy estimation approach presented in this paper provides an effective and flexible procedure for reconciling micro data derived from a household survey with macro data derived from a Social Accounting Matrix or national accounts. While the method suffices for our main objective (reconciling data from macro and micro sources), it can certainly be improved by adding more information. The flexibility of the method allows adding information derived from many different types of sources.

While this procedure has been developed to support microsimulation modeling, other applications can be considered. For example, reconciling household and production surveys with information gathered at the regional level in an economy can provide an efficient approach to estimating a SAM with extensive regional and household dissagregation.

Possible extensions of the procedure in the context of household surveys include simultaneous estimation of household relationships, use of other data, and specifying "errors in variables" to incorporate survey data errors. Such extensions have been used in other contexts, and do considerably increase the size of the estimation problem.

References

Brooke, A., D. Kendrick, and A. Meeraus, *GAMS, A User's Guide*, The Scientific Press, San Francisco, 1998.

Cogneau, D. and A. S. Robilliard, "Income Distribution, Poverty, and Growth in Madagascar: Microsimulations in General Equilibrium Framework," Paper presented at the Forty-Eighth International Atlantic Economic Conference, October 7–10, Montreal, Canada, 1999.

Dirkse, S. and M. Ferris, "The PATH Solver: A Non Monotone Stabilization Scheme for Computing Stationary Points," *Optimization Methods and Software*, 5, 125–56, 1995.

Golan, A., "Maximum Entropy, Likelihood and Uncertainty: A Comparison," in G. J. Erickson, J. T. Rychert, and C. R. Smith (eds), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1998.

Golan, A., G. Judge, and D. Miller, *Maximum Entropy Econometrics, Robust Estimation with Limited Data*, John Wiley & Sons, New York, 1996.

Jaynes, E. T., "Information Theory and Statistical Mechanics," *Physics Review*, 106, 620–30, 1957.

Kapur, J. N. and H. K. Kesavan, *Entropy Optimization Principles with Applications*, Academic Press, San Diego, 1992.

Kullback, S. and R. A. Leibler, "On Information and Sufficiency," *Ann. Math. Stat*. 4, 99–111, 1951.

Magnus, J. R., J. W. van Tongeren, and A. F. de Vos, "National Accounts Estimation using Indicator Ratios," *The Review of Income and Wealth*, 46(3), 329–50, 2000.

Pyatt, G., "SAMs, the SNA and National Accounting Capabilities," *The Review of Income and Wealth*, 37(2), 177–98, 1991.

Razafindrakoto M. and F. Roubaud, "Une Matrice de Comptabilité Sociale pour Madagascar," Working Paper MADIO no. 9744/E, Antananarivo, 1997.

Robinson, S., A. Cattaneo, and M. El-Said, "Estimating a Social Accounting Matrix Using Cross Entropy Methods," *Economic Systems Research*, 13(1), 2001.

Rutherford, T., "Extension of GAMS for Complementarity Problems Arising in Applied Economic Analysis," *Journal of Economic Dynamics and Control*, 19, 1299–324, 1995.

Shannon, C. E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 379–423, 1948.
Theil, H., *Economics and Information Theory*, Rand McNally, New York, 1967.
Van Tongeren, J. W., "Development of an Algorithm for the Compilation of National Accounts and Related Systems of Statistics," *The Review of Income and Wealth*, 32(1), 25–47, 1986.
Zellner, A., "Optimal Information Processing and Bayes Theorem," *American Statistician*, 42, 278–84, 1988.
———, "Bayesian Methods and Entropy in Economics and Econometrics," in W. T. Grandy and L. H. Shick (eds), *Maximum Entropy and Bayesian Methods*, Kluwer, Dordrecht, 17–31, 1990.