TMD DISCUSSION PAPER NO. 50

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

**Anne-Sophie Robilliard**

**Sherman Robinson**

**International Food Policy Research Institute**

**Trade and Macroeconomics Division**

**International Food Policy Research Institute**

**2033 K Street, N.W.**

**Washington, D.C. 20006, U.S.A.**

**November 1999**

**(Revised January 2001)**

***TMD Discussion Papers contain preliminary material and research results, and are circulated prior to a full peer review in order to stimulate discussion and critical comment. It is expected that most Discussion Papers will eventually be published in some other form, and that their content may also be revised. This paper is available at http://www.cgiar.org/ifpri/divs/tmd/dp/papers/dp50.pdf***

## Abstract

This paper presents 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, but the macro aggregates may be measured with error. The economic data gathered in the household survey are assumed to be accurate, or have been adjusted to be accurate. 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 estimation approach represents an efficient "information processing rule" using 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 using a cross-entropy metric 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.



## Acknowledgments

The authors wish to thank participants at seminars at Yale University and IFPRI for helpful comments on earlier drafts of this paper.

**Table of Contents**

## 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 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 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 dissagregate the national data to incorporate greater sectoral, regional, or household detail (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.

The next section presents the background and a mathematical description of the estimation problem, while the following section present an application to the case of Madagascar.

## 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 households 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

[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.

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" 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 $\overline{p}_k$ 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:

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

subject to moment consistency constraints

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

and the adding-up normalization constraint

$$\sum_{k=1}^{K} p_k = 1 \tag{3}$$

where $\{y_1, y_2, ..., y_T\}$ 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) $\{p_1, p_2, ..., p_K\}$. 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, 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:

$$L = \sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{\overline{p}_k}\right) + \sum_{t=1}^{T} l_t \left( y_t - \sum_{k=1}^{K} p_k f_t(x_k) \right) + m\left(1 - \sum_{k=1}^{K} p_k\right) \qquad (4)$$

The first-order conditions are:

$$\frac{\partial L}{\partial p_k} = \ln p_k - \ln \overline{p}_k + 1 - \sum_{t=1}^{T} l_t f_t(x_k) - m = 0, \qquad k \in [1,...,K] \qquad (5)$$

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

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

$$\frac{\partial L}{\partial m} = 1 - \sum_{k=1}^{K} p_k = 0 \qquad (7)$$

The solution can be written as:

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

where

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

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 all the constraints were not 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:

$$e_t = \sum_{l} w_{t,l} \overline{v}_{t,l} \qquad (10)$$

where $e_t$ = error value

$w_{t,l}$ = error weights estimated in the CE procedure

$\overline{v}_{t,l}$ = error support set

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

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:

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

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

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

With errors in aggregates, the constraints of the problem are

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

and additional adding-up constraints on the error weights

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

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 $s$ , we used a support set with five terms equal to $(-3s, -s, 0, s, 3s)$. 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 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.

## 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).

### *Data Sources*

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

[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. See Appendix A for the details of the computation.

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 others 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*

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 income 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. Finally, 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. Again, this choice is made by default, because of lack of data. 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 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 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 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 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 consistent 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, 1988), using a mixed complementarity formulation (Rutherford, 1995 and Dirkse and Ferris, 1995). 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, following the mathematical description of the estimation procedure.

Using the MCP formulation requires writing the first order conditions of the optimization problem, yielding a square equation system that explicitly includes shadow-price variables and complementary slackness conditions. The resulting problem is relatively large, with more than 4,500 equations and variables. The results using this approach are identical to those from a classic optimization formulation. The MCP approach, however, performs much better than a nonlinear programming algorithm.[5]

*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 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.

Three simulations are presented. 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) [6]. 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),

---

[5] A closely-related solution approach is to derive the dual programming problem, which is straightforward when the constraints are all linear, and solve it using a standard NLP algorithm. Golan, Judge, and Miller (1996) report success with this method. The approach is similar to the MCP approach in that both take advantage of the fact that there are far fewer shadow prices in the dual than endogenous variables in the primal.

[6] 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.

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[7]. As a result, the new samples are smaller.

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 income shares per quintile (Table 4) show that relative income distribution does not change dramatically in the first reweighted sample. However, Gini and Theil indexes (Tables 5 and 6, column FOM) show more sensitivity to the reweighing procedures. This led us to introduce higher order moment constraints on income. The result is actually a "tighter" income distribution in the SOM simulation but the introduction of errors in aggregates does not change the results in terms of income distribution. The decomposition of the Gini index in Table 5 shows how different income sources and their distribution are affected by the reweighting procedures[8]. In the prior sample, agricultural, informal, and formal contribute in increasing order both to income specific Gini indexes and to total inequality. The relative contributions appear to differ significatively in the reweighted samples, especially for informal income.

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

[8] The decomposition of income inequality by source of income allows measurement of the contribution of the different sources of income to overall income inequality and can be used to determine whether any particular source of income contributes to increase or decrease income inequality (Sadoulet and De Janvry, 1995).

We finally present lambda values associated with constraints imposed on the problem. The bigger the absolute value, the more binding is the constraint. In general, big lambda values point out to constraints that should be looked at more carefully. In our case, the income aggregation constraint appears to be the most important (Table 7). Note that all the lambdas decrease with the introduction of errors in aggregates, since this specification “loosens up” all the constraints (see Table 3).

## 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 (1998). *GAMS a User's Guide*. The Scientific Press, San Francisco.

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

Dirkse, S. and M. Ferris (1995). "The PATH solver: A non monotone stabilization scheme for computing stationary points." *Optimization Methods and Software*, 5, 125-56.

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

Golan, A., G. Judge, and D. Miller (1996). *Maximum Entropy Econometrics, Robust Estimation with Limited* .

Jaynes, E.T. (1957). "Information theory and statistical mechanics." *Physics Review* 106, 620-30.

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

Kullback, S. and R. A. Leibler (1951). "On Information and Sufficiency." Ann. Math. Stat. 4, 99-111.

Pyatt, G. (1991). "SAMs, the SNA and National Accounting Capabilities." *Review of Income and Wealth* 37(2) 177-198.

Razafindrakoto M., and F. Roubaud (1997). "Une Matrice de Comptabilité Sociale pour Madagascar." DT MADIO n°9744/E, Antananarivo.

Robinson, S., A. Cattaneo, and M. El-Said (20001). "Estimating a Social Accounting Matrix Using Cross Entropy Methods." Forthcoming, *Economic Systems Research*. Also available as TMD Discussion Paper No. 33.

Rutherford, T. (1995). "Extension of GAMS for complementarity problems arising in applied economic analysis." *Journal of Economic Dynamics and Control* 19 (1995) 1299-1324.

Shannon, C. E. (1948). “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27, 379-423.

Theil, H. (1967). *Economics and Information Theory*. Rand McNally.

Tongeren, J.W. van (1986). “Development of an algorithm for the compilation of national accounts and related systems of statistics”. *Review of Income and Wealth* 32(1) 25-47.

Zellner, A. (1990). “Bayesian Methods and Entropy in Economics and Econometrics”. In W. T. Grandy and L. H. Shick (Eds.), *Maximum Entropy and Bayesian Methods*, pp. 17-31. Kluwer, Dordrecht.

Zellner, A. (1988). “Optimal Information Processing and Bayes Theorem”. *American Statistician* 42, 278-84.

**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 |

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

**Table 2: New Weights Distribution**

| | Prior | FOM | SOM | EIA |
|---|---|---|---|---|
| Mean Weight | 557.0 | 593.7 | 593.7 | 593.7 |
| Standard Deviation | 365.7 | 407.2 | 413.3 | 389.0 |
| Maximum Weight | 1,901 | 3668 | 4061 | 4008 |
| Minimum Weight | 114 | 0 | 0 | 0 |
| Number of zero weights | 0 | 165 | 163 | 90 |

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

**Table 3: Selected Aggregate Results**

| | Adjusted Prior | FOM | SOM | EIA |
|---|---|---|---|---|
| Total Number of Households ('000)* | 2,649 | 2,649 | 2,649 | 2,649 |
| Total Population ('000)* | 13,059 | 13,059 | 13,059 | 13,037 |
| Total Income (millions of 95 FMG)* | 11,315 | 11,315 | 11,315 | 11,509 |
| Mean Per Capita Income ('000 of 95 FMG)* | 866 | 866 | 866 | 883 |
| Share Rural Population (%)* | 82.8 | 75.2 | 75.2 | 76.2 |
| Share Males (%) | 49.5 | 49.8 | 49.7 | 49.6 |
| Mean Age (years) | 21.5 | 21.4 | 21.4 | 21.5 |

* used as constraints in the program

**Table 4: Income distribution**

| | Prior | FOM | SOM | EIA |
|---|---|---|---|---|
| Income Share of the 10% poorest | 1.7 | 1.7 | 1.7 | 1.6 |
| Income Share of the 20% poorest | 4.6 | 4.5 | 4.4 | 4.4 |
| Income Share of the 40% poorest | 12.7 | 12.4 | 12.3 | 12.3 |
| Income Share of the 10% richest | 43.2 | 45.6 | 45.5 | 45.7 |

**Table 5: Contribution of different income sources to inequality**

| | Prior | FOM | SOM | EIA |
|---|---|---|---|---|
| Overall Gini Index | 53.7 | 54.8 | 54.5 | 54.8 |
| | | | | |
| Gini Index for Agricultural Income | 62.0 | 63.8 | 64.8 | 64.7 |
| Share of Agricultural Income (%) | 33.4 | 35.0 | 35.4 | 34.4 |
| Contribution to Overall Gini (%) | 20.5 | 22.6 | 23.7 | 22.5 |
| | | | | |
| Gini Index for Informal Income | 72.9 | 65.9 | 65.8 | 66.0 |
| Share of Informal Income (%) | 30.4 | 18.2 | 18.3 | 17.9 |
| Contribution to Overall Gini (%) | 30.1 | 10.9 | 10.9 | 10.7 |
| | | | | |
| Gini Index for Formal Income | 86.1 | 87.4 | 86.8 | 86.1 |
| Share of Formal Income (%) | 36.1 | 46.8 | 46.3 | 47.7 |
| Contribution to Overall Gini (%) | 49.4 | 66.6 | 65.4 | 66.8 |

**Table 6: Decomposition of the Theil index**

| | Prior | FOM | SOM | EIA |
|---|---|---|---|---|
| Theil Index | 63.8 | 71.8 | 69.4 | 69.5 |
| - between | 7.8 | 8.8 | 6.9 | 7.7 |
| - within | 56.0 | 63.1 | 62.5 | 61.9 |
| Theil Index for Urban Area | 69.1 | 82.4 | 77.3 | 77.0 |
| Theil Index for Rural Area | 49.5 | 47.8 | 51.9 | 51.0 |

**Table 7: Lambda Values for Constraints**

| | FOM | SOM | EIA |
|---|---|---|---|
| Total population | 0.08 | 0.08 | 0.03 |
| Total income | -1.98 | -1.99 | -0.36 |
| Agricultural income share | 0.75 | 0.73 | 0.14 |
| Informal income share | 0.19 | 0.19 | -0.01 |
| Formal labor income share | 0.40 | 0.40 | 0.08 |
| Formal capital income share | 0.45 | 0.45 | 0.08 |
| Share rural population | -0.62 | -0.63 | -0.27 |
| Total rural income (second moment of log) | 0.00 | 0.02 | 0.00 |
| Total urban income (second moment of log) | 0.00 | 0.00 | 0.00 |

Notes: i) since all the constraints have been normalized to one, lambda values are comparable; ii) zero means less than $10^{-3}$.

## Appendix A: Case of a five-weight error distribution

For the case of five-parameter error distribution, there are five weights, $\bar{w}$, to be estimated—the set l consists of five elements. That is we are incorporating more information about and error distribution; more moments, including the variance, skewness, and kurtosis. Assuming a prior mean of zero and a prior value of kurtosis consistent with a prior normal distribution with mean zero, variance $s^2$, and kurtosis equal to $3s^4$. In this case, the prior weights, $\bar{w}$, are specified so that:

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

and

$$\sum_l \bar{w}_{i,l} \cdot \bar{v}_{i,l}^4 = 3s^4$$

The prior weights and support set are also symmetric, so the prior on all odd moments is zero. The choice of $\pm 1$ standard error for $\bar{v}_{i,2}$ and $\bar{v}_{i,4}$ is arbitrary and the actual moments are estimated as part of the estimation procedure. in this case we get:

$$\bar{v}_{i,1} = -3s$$
$$\bar{v}_{i,2} = -s$$
$$\bar{v}_{i,3} = 0$$
$$\bar{v}_{i,4} = +s$$
$$\bar{v}_{i,5} = +3s$$

Since

$$s^2 = \bar{w}_{i,1} \cdot \left(9s^2\right) + \bar{w}_{i,2} \cdot \left(s^2\right) + \bar{w}_{i,3} \cdot (0) + \bar{w}_{i,2} \cdot \left(s^2\right) + \bar{w}_{i,1} \cdot \left(9s^2\right)$$
$$3s^4 = \bar{w}_{i,1} \cdot \left(81s^4\right) + \bar{w}_{i,2} \cdot \left(s^4\right) + \bar{w}_{i,3} \cdot (0) + \bar{w}_{i,2} \cdot \left(s^4\right) + \bar{w}_{i,1} \cdot \left(81s^4\right)$$

or

$$18\bar{w}_{i,1} + 2\bar{w}_{i,2} = 1$$
$$162\bar{w}_{i,1} + 2\bar{w}_{i,2} = 3$$

solving for the $\bar{w}$s we get

$$\bar{w}_{i,1} = \frac{1}{72} \quad ; \quad \bar{w}_{i,2} = \frac{27}{72} \quad ; \quad \bar{w}_{i,3} = \frac{16}{72} \quad ; \quad \bar{w}_{i,4} = \frac{27}{72} \quad ; \quad \bar{w}_{i,5} = \frac{1}{72}$$

**Appendix B: GAMS code**

What follows is a listing of the GAMS program used to implement the cross entropy estimation approach. A quick list of some GAMS features are listed below:

- Five principal self-explanatory keywords define the nature of the elements declared: “SETS”, “PARAMETERS”, “VARIABLES”, “EQUATION”, “MODEL”;
- Four suffixes can be linked to variables:
  - “.FX’ indicates a fixed variables (treated as a constant),
  - “.L” indicates the level or solution value of a variable,
  - “.LO” and “.UP” indicate the lower and upper bounds of a variable;
- the symbol “$” introduces a conditional statement;
- an asterisk in the first column indicates a comment;
- an “ALIAS” statement is used to give another name to a previously declared set;
- the “$libinclude xlimport” statement is used to import data from an Excel file;
- the “$libinclude xlexport” statement is used to export data to an Excel file.

For additional information about the GAMS syntax, see Brooke, Kendrick, and Meeraus (1988).

```
$TITLE MICMAC Estimate household population weights using survey
information

* Programmed by: Anne-Sophie Robilliard   a.s.robilliard@cgiar.org
*                Sherman Robinson         s.robinson@cgiar.rog
*
* Trade and Macroeconomics Division
* International Food Policy Research Institute (IFPRI)
* 2033 K St., NW
* Washington, DC 20006
*
* April 2000
*

$OFFSYMLIST OFFSYMXREF OFFUPPER
*OFFLISTING

sets
  k     Households                                          /h1*h4458/

  t     Moment constraints on household weights             /c1*c13/

  l     Set for errors in aggregates                        /l1*l5/

  var   /
        id
        mil
        taille
        sexe
        age
        sup
        deptot
        poids0
        poids1
        revtot0
        revtot2
        revtot3
        shagr
        shcvg
        shinf
        shfor
        shdiv
        shexo
        /;

alias(k,kp);
```

```
parameter micdat(k,var);
$libinclude xlimport micdat micdat.xls A1..S4459
*$include micdat.inc
display micdat;

scalars
            popgrw   population growth rate          /1.029/

*as recovering weights and population for discarded observations
            dischh   number of discarded households  /18917/
            discpop  number of discarded inhabitants /92573/
;

Parameters
  hhmil(k)
  hhsize(k)
  hhweight(k)
  hhinc(k)
  hhincagr(k)
  hhinccvg(k)
  hhincinf(k)
  hhincfor(k)
  hhincdiv(k)
  NHHTOT0
  POPTOT0
  INCTOT0
  INCTOT95
  INCAGR95
  INCCVG95
  INCINF95
  INCFOR95
  INCDIV95
  NHHTOT95
  POPTOT95
  delta
  adjfac1
;

parameter sh(*)
 /
 AGR95  0.363
 CVG95  0.190
 INF95  0.174
 FOR95  0.194
 DIV95  0.225
```

```
 /;

 INCTOT95     = 11.315*1e9;

 INCAGR95     = sh('AGR95')*INCTOT95;
 INCCVG95     = sh('CVG95')*INCAGR95;
 INCINF95     = sh('INF95')*INCTOT95;
 INCFOR95     = sh('FOR95')*INCTOT95;
 INCDIV95     = sh('DIV95')*INCTOT95;

 hhmil(k)     = micdat(k,'mil');
 hhweight(k)  = micdat(k,'poids0');
 hhsize(k)    = micdat(k,'taille');
 hhinc(k)     = micdat(k,'revtot3');

 NHHTOT0      = sum(k, hhweight(k));
 POPTOT0      = sum(k, hhsize(k)*hhweight(k));
 INCTOT0      = sum(k, hhinc(k)*hhweight(k));

 NHHTOT95     = (NHHTOT0 + dischh)*popgrw**2 ;
 POPTOT95     = (POPTOT0 + discpop)*popgrw**2 ;

*ASR homothetic adjustment of mean income to match macro aggregate
 adjfac1      = (NHHTOT0/NHHTOT95)*(INCTOT95/INCTOT0);
display adjfac1;

 hhinc(k)     = adjfac1*hhinc(k);

 hhincagr(k)  = hhinc(k)*micdat(k,'shagr');
 hhincagr(k)  = hhinc(k)*micdat(k,'shagr');
 hhincinf(k)  = hhinc(k)*micdat(k,'shinf');
 hhincfor(k)  = hhinc(k)*micdat(k,'shfor');
 hhincdiv(k)  = hhinc(k)*micdat(k,'shdiv');

 hhinccvg(k)  = hhincagr(k)*micdat(k,'shcvg');

 delta        = 1e-5;

display
  NHHTOT0
  POPTOT0
  INCTOT0
  INCTOT95
  INCAGR95
  INCCVG95
  INCINF95
```

```
  INCFOR95
  INCDIV95
  NHHTOT95
  POPTOT95
;

*ASR/SR Define moment constraints. Constraint zero is sum of weights
equals 1

Parameter XBAR(t,k)   Household data
          Y0(t)       Unscaled Moment values
          Y(t)        Scaled Moment values
          PUNI(k)     Uniform Prior household weights
          PBAR(k)     Prior household weights
          VBAR(t,l)   Support set for errors in aggregates
          W0(t,l)   Prior weights for errors in aggregates
          SIGMAY(t)   Prior on standard error of aggregate error
          alpha       Weight of "errors in aggregates" in maximand
          sigp        Sum of Ps
          sigw(t)     Sums of Ws;

$ontext
*ASR/SR Fill in data from households to define moment constraints
*       Constraint c0 is the adding up to one constraint

 XBAR("c0",k) = 1 ;
 Y0("c0")     = 1 ;
$offtext

* Total population
 XBAR("c1",k) = hhsize(k) ;
 y0("c1")     = SUM(k, XBAR("c1",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Total income
 XBAR("c2",k) = hhinc(k) ;
 y0("c2")     = SUM(k, XBAR("c2",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Agricultural income
 XBAR("c3",k) = hhincagr(k) ;
 y0("c3")     = SUM(k, XBAR("c3",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Informal income
 XBAR("c4",k) = hhincinf(k) ;
```

```
 y0("c4")     = SUM(k, XBAR("c4",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Formal labor income
 XBAR("c5",k) = hhincfor(k) ;
 y0("c5")     = SUM(k, XBAR("c5",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Formal capital income
 XBAR("c6",k) = hhincdiv(k) ;
 y0("c6")     = SUM(k, XBAR("c6",k)*hhweight(k))/SUM(kp, hhweight(kp))
;

* Cash crop income
 XBAR("c7",k) = hhinccvg(k);
 y0("c7")     = SUM(k, XBAR("c7",k)*hhweight(k))/SUM(kp,
hhweight(kp));

*ASR/SR  Share rural population
 XBAR("c8",k) = hhmil(k);
 y0("c8")     = SUM(k, XBAR("c8",k)*hhweight(k))/SUM(kp,
hhweight(kp));

*ASR/SR  Total rural income (first moment of log)
*XBAR("c9",k)  = hhmil(k)*log(hhinc(k));
 XBAR("c9",k)  = hhmil(k)*hhinc(k);
 y0("c9")      = SUM(k, XBAR("c9",k)*hhweight(k))
                 /SUM(kp, hhmil(kp)*hhweight(kp));
*                /SUM(kp, hhmil(kp)*hhweight(kp)*hhsize(kp));
*                /SUM(kp, hhweight(kp));

*ASR/SR  Total urban income (first moment of log)
*XBAR("c10",k) = (1-hhmil(k))*log(hhinc(k));
 XBAR("c10",k) = (1-hhmil(k))*hhinc(k);
 y0("c10")     = SUM(k, XBAR("c10",k)*hhweight(k))
                 /SUM(kp, (1-hhmil(kp))*hhweight(kp));
*                /SUM(kp, (1-hhmil(kp))*hhweight(kp)*hhsize(kp));
*                /SUM(kp, hhweight(kp));

*ASR/SR  Total rural income (second moment of log)
 XBAR("c11",k) = hhmil(k)*log(hhinc(k))**2;
 y0("c11")     = SUM(k, XBAR("c11",k)*hhweight(k))
                 /SUM(kp, hhmil(kp)*hhweight(kp));
*                /SUM(kp, hhweight(kp));

*ASR/SR  Total urban income (second moment of log)
 XBAR("c12",k) = (1-hhmil(k))*log(hhinc(k))**2;
 y0("c12")     = SUM(k, XBAR("c12",k)*hhweight(k))
                 /SUM(kp, (1-hhmil(kp))*hhweight(kp));
*                /SUM(kp, hhweight(kp));

display Y0 ;

parameters shrur0, shrur1;
*shrur0 = sum(k, hhmil(k)*hhweight(k)*hhsize(k))/sum(kp,
hhweight(kp)*hhsize(kp));
 shrur0 = sum(k, hhmil(k)*hhweight(k))/sum(kp, hhweight(kp));
 shrur1 = 0.75;

*imposing aggregate moments
 Y0("c1")     = POPTOT95/NHHTOT95;
 Y0("c2")     = INCTOT95/NHHTOT95;
 Y0("c3")     = INCAGR95/NHHTOT95;
 Y0("c4")     = INCINF95/NHHTOT95;
 Y0("c5")     = INCFOR95/NHHTOT95;
 Y0("c6")     = INCDIV95/NHHTOT95;
 Y0("c7")     = INCCVG95/NHHTOT95;
 Y0("c8")     = shrur1;
 Y0("c9")     = Y0("c9")*Y0("c8");
 Y0("c10")    = Y0("c10")*(1-Y0("c8"));
 Y0("c11")    = Y0("c11")*Y0("c8");
 Y0("c12")    = Y0("c12")*(1-Y0("c8"));

*as the overall mean income is equal to the weighted sum of the mean
rural income
* and the mean urban income. mean incomes from different origins need
to be
* reinitialized once the urban and the rural shares as well as the
first
* and second moments of rural and urban income are fixed
*Y0("c3")     = sh('AGR95')*(Y0("c9")+Y0("c10"));
*Y0("c4")     = sh('INF95')*(Y0("c9")+Y0("c10"));
*Y0("c5")     = sh('FOR95')*(Y0("c9")+Y0("c10"));
*Y0("c6")     = sh('DIV95')*(Y0("c9")+Y0("c10"));
*Y0("c7")     = sh('CVG95')*sh('AGR95')*(Y0("c9")+Y0("c10"));

display Y0 ;

*Normalize moments to all equal one.
*Constraint "c0" is already normalized.

 XBAR(t,k)$Y0(t) = XBAR(t,k)/Y0(t) ;
```

```
 Y(t)            = 1 ;

 sigmay(t)       = .15 ;

*ASR/SR  Ratio urban to rural income
 y0("c13")     = SUM(k, hhweight(k)*hhmil(k)*hhinc(k))
                 /SUM(k, hhweight(k)*(1-hhmil(k))*hhinc(k));
 XBAR("c13",k) = (hhmil(k)-y0("c13")*(1-hhmil(k)))*hhinc(k);

display Y ;

 SET
      at(t)   Active Moment constraints on household weights
            /
*           c0          Sum to one
            c1          Total population
            c2          Total income
            c3          Agricultural income
            c4          Informal income
            c5          Formal labor income
            c6          Formal capital income
*           c7          Cash crop income
            c8          Share rural population
*           c9          Rural income (first moment of log)
*           c10         Urban income (first moment of log)
            c11         Rural income (second moment of log)
            c12         Urban income (second moment of log)
*           c13         Ratio rural to urban income
            /
;

*Define prior on household weights
 PUNI(k)     = 1/4458;
 PBAR(k)     = hhweight(k)/SUM(kp, hhweight(kp));
*PBAR(k)     = PUNI(k);

*Set errors in aggregates support set and share in maximand

* VBAR(t,'l1') = -1;
* VBAR(t,'l2') = 0;
* VBAR(t,'l3') = 1;
$ontext
*Set constants for three parameter error distribution
  VBAR(t,'l1') = -3*sigmay(t);
  VBAR(t,'l2') = 0;
  VBAR(t,'l3') = +3*sigmay(t);

  W0(t,"l1") =  1/18 ;
  W0(t,"l2") = 16/18 ;
  W0(t,"l3") =  1/18 ;
$offtext

*Set constants for five parameter error distribution
  VBAR(t,'l1') = -3*sigmay(t);
  VBAR(t,'l2') = -1*sigmay(t);
  VBAR(t,'l3') = 0;
  VBAR(t,'l4') = +1*sigmay(t);
  VBAR(t,'l5') = +3*sigmay(t);

  W0(t,"l1") = .01389 ;
  W0(t,"l2") = .375 ;
  W0(t,"l3") = .22222 ;
  W0(t,"l4") = .375 ;
  W0(t,"l5") = .01389 ;

 VARIABLES
  P(k)          Household population weights
  LAMBDA(t)     Lagrangian multiplier on summing-up constraints
  MU            Lagrangian multiplier on additivity constraint
  W(t,l)        Errors in aggregates weights
  GAMMA(t)      Lagrangian multiplier on additivity constraint
  DENTROPY      Cross entropy minimand
 ;

 P.L(k)        = PBAR(k) ;
 LAMBDA.L(at)  = 1 ;
 W.L(at,l)     = W0(at,l);
 GAMMA.L(at)   = 1 ;
 DENTROPY.L    = 0 ;

EQUATIONS
 FOCP(k)       First order condition w.r.t P
 FOCW(t,l)     First order condition w.r.t W
 MOMENT(t)     Moment constraints
 ADDP          Additivity constraint for P
 ADDW(t)       Additivity constraint for W
 ENTROPY       Entropy difference definition
 ;

 FOCP(k)..       (1-alpha)*( LOG((P(k)+delta)/(PBAR(k)+delta))
                             + P(k)*(PBAR(k)+delta)/(P(k)+delta) )
                 - sum(at, LAMBDA(at)* XBAR(at,k)) - MU =G= 0 ;
```

```
 FOCW(at,l)..    alpha*( LOG((W(at,l)+delta)/(W0(at,l)+delta))
                         + W(at,l)*(W0(at,l)+delta)/(W(at,l)+delta) )
                 + LAMBDA(at)* VBAR(at,l) - GAMMA(at) =G= 0 ;

 MOMENT(at)..    sum(k, P(k)*XBAR(at,k)) =E= y(at) + sum(l,
W(at,l)*VBAR(at,l)) ;

 ADDP..          sum(k, P(k)) =E= sigp ;

 ADDW(at)..      sum(l, W(at,l)) =E= sigw(at) ;

 ENTROPY..       DENTROPY
                 =E= (1-alpha)*sum(k,
P(k)*LOG((P(k)+delta)/(PBAR(k)+delta)))
                     + alpha*sum((at,l),
W(at,l)*LOG((W(at,l)+delta)/(W0(at,l)+delta))) ;

 P.LO(k)       = 0 ;
 W.LO(at,l)    = 0 ;

*P.LO(k)       = 100/SUM(kp, hhweight(kp)) ;
*P.UP(k)       = 5000/SUM(kp, hhweight(kp)) ;

*parameter checkMCF(at) ;
*checkMCF(at) = sum(k, XBAR(at,k)*P.up(k)) ;
*display checkMCF ;

 OPTION LIMROW = 1000, LIMCOL = 0, ITERLIM = 100000, RESLIM = 50000.0
;
 OPTION SOLPRINT = ON ;
 OPTION MCP = PATH ;

 MODEL MICMAC /
   FOCP.P
   FOCW.W
   MOMENT.LAMBDA
   ADDP.MU
   ADDW.GAMMA
   ENTROPY.DENTROPY
 / ;

 MICMAC.holdfixed    = 1 ;

SET sim  /
        INI      Initial
        ADJ      Adjusted
        OBJ      Objective
        MICMAC5  Min. cross-ent. s.t. 1st order moments
        MICMAC7  Min. cross-ent. s.t. 1st & 2nd order moments
        MICMAC9  Min. cross-ent. s.t. 1st & 2nd order moments w.
errors in aggregates
        /
;
display sim;

parameters
 res(k,*)
 poids(k,*)
 NE(sim) Normalized entropy for sample weights only
 macres(*,sim)
 hhwelf(k), MWELF, MWELFN, MWELFU, MWELFUN, MWELFR, MWELFRN, hhwuc(k),
hhwucn(k)
 incres(*,sim)
 tabres(*,sim)
 shres(*,sim)
 zeros(k,sim)
 nzeros(sim)
 carzeros(*,sim)
 lbdres(t,sim)
 mures(sim)
 gamres(t,sim)
 wres(sim,t,l)
 errors(t,sim)  errors in aggregates (precent deviation to target
values)
;

poids(k,'id')  = micdat(k,'id');
poids(k,'ini') = PBAR(k)*NHHTOT0;
nzeros(sim)    = 0$(ord(sim) le 3);

LOOP(sim$(ord(sim) gt 3),

 IF(ORD(sim) EQ 4,
            at('c11')  = NO;
            at('c12')  = NO;
            alpha      = 0;
            W.l(at,l)  = 0;
            sigp       = 1;
            sigw(at)   = 0;
   );
```

```
 IF(ORD(sim) EQ 5,
           at('c11')  = YES;
           at('c12')  = YES;
           alpha      = 0;
           W.l(at,l)  = 0;
           sigp       = 1;
           sigw(at)   = 0;
   );

 IF(ORD(sim) EQ 6,
           at('c11')  = YES;
           at('c12')  = YES;
           alpha      = 0.5;
           W.l(at,l)  = W0(at,l);
           sigp       = 1;
           sigw(at)   = 1;
   );

 SOLVE MICMAC USING MCP ;

 res(k,'poids1')       = P.L(k)*NHHTOT95;
 poids(k,sim)          = P.L(k)*NHHTOT95;

$include simres.inc
);

 macres('WGHTOT','ini')  = sum(k,hhweight(k))/1e3;
 macres('WGHTOT','adj')  = NHHTOT0/1e3;
 macres('WGHTOT','obj')  = NHHTOT95/1e3;

 macres('POPTOT','ini')  = sum(k,hhweight(k)*hhsize(k))/1e3;
 macres('POPTOT','adj')  = POPTOT0/1e3;
 macres('POPTOT','obj')  = POPTOT95/1e3;

 macres('INCTOT','ini')  = INCTOT0/1e6;
 macres('INCTOT','adj')  = sum(k,hhweight(k)*hhinc(k))/1e6;
 macres('INCTOT','obj')  = INCTOT95/1e6;

 macres('MEANPCI','ini') = INCTOT0/POPTOT0;
 macres('MEANPCI','adj') = sum(k,hhweight(k)*hhinc(k))
                          /sum(k,hhweight(k)*hhsize(k));
 macres('MEANPCI','obj') = INCTOT95/POPTOT95;


 incres('MEANINC','ini') = sum(k,hhweight(k)*hhinc(k))
                          /sum(k,hhweight(k));
 incres('MINCRUR','ini') = sum(k,hhmil(k)*hhweight(k)*hhinc(k))
                          /sum(k,hhmil(k)*hhweight(k));
 incres('MINCURB','ini') = sum(k,(1-hhmil(k))*hhweight(k)*hhinc(k))
                          /sum(k,(1-hhmil(k))*hhweight(k));
 incres('MPCIRUR','ini') = sum(k,hhmil(k)*hhweight(k)*hhinc(k))
                          /sum(k,hhmil(k)*hhweight(k)*hhsize(k));
 incres('MPCIURB','ini') = sum(k,(1-hhmil(k))*hhweight(k)*hhinc(k))
                          /sum(k,(1-hhmil(k))*hhweight(k)*hhsize(k));
 incres('THEIL','ini')   = 100*sum(k,hhwuc(k)*(hhwelf(k)/MWELF)
                                *log(hhwelf(k)/MWELF))
                          /sum(k,hhwuc(k));


 tabres('SUPTOT','ini')  = sum(k,hhweight(k)*micdat(k,'sup'))/1e3;
 tabres('MEANSUP','ini') = sum(k,hhweight(k)*micdat(k,'sup'))
                          /sum(k,hhweight(k));
 tabres('SHMALE','ini')  =
sum(k,hhweight(k)*micdat(k,'sexe')*hhsize(k))
                          /sum(k,hhweight(k)*hhsize(k));
 tabres('SHRURAL','ini') =
sum(k,hhweight(k)*micdat(k,'mil')*hhsize(k))
                          /sum(k,hhweight(k)*hhsize(k));
 tabres('MEANAGE','ini') =
sum(k,hhweight(k)*micdat(k,'age')*hhsize(k))
                          /sum(k,hhweight(k)*hhsize(k));


 shres(' shagr','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shagr')*hhinc(k))
                          /sum(k,hhweight(k)*hhinc(k));
 shres(' shagr','obj')   = 100*INCAGR95/INCTOT95;

 shres(' shinf','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shinf')*hhinc(k))
                          /sum(k,hhweight(k)*hhinc(k));
 shres(' shinf','obj')   = 100*INCINF95/INCTOT95;

 shres(' shfor','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shfor')*hhinc(k))
                          /sum(k,hhweight(k)*hhinc(k));
 shres(' shfor','obj')   = 100*INCFOR95/INCTOT95;

 shres(' shdiv','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shdiv')*hhinc(k))
                          /sum(k,hhweight(k)*hhinc(k));
 shres(' shdiv','obj')   = 100*INCDIV95/INCTOT95;
```

```
 shres(' shexo','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shexo')*hhinc(k))
                           /sum(k,hhweight(k)*hhinc(k));
 shres(' shexo','obj')   = 100 - shres(' shagr','obj') - shres('
shinf','obj')
                               - shres(' shfor','obj') - shres('
shdiv','obj');

 shres(' shcvg','ini')   =
100*sum(k,hhweight(k)*micdat(k,'shcvg')*hhincagr(k))
                           /sum(k,hhweight(k)*hhincagr(k));
 shres(' shcvg','obj')   = 100*INCCVG95/INCAGR95;

 res(k,'id')           = micdat(k,'id');
 res(k,'adj')$hhinc(k) = adjfac1;
 option decimals=2;
*display res;
*$libinclude xlexport res micmac9.xls A1..D4459

*$libinclude xlexport poids micmac.xls A1..F4459

option decimals=4;
 display NE;
option decimals=2;
 display lbdres, mures, gamres, wres, errors;

option decimals=0;
 display macres, incres;
option decimals=3;
 display tabres;
option decimals=1;
 display shres;
 display nzeros;
 display carzeros;
```

# LIST OF DISCUSSION PAPERS

No. 1- "Land, Water, and Agriculture in Egypt: The Economywide Impact of Policy Reform" by Sherman Robinson and Clemen Gehlhar (January 1995)

No. 2- "Price Competitiveness and Variability in Egyptian Cotton: Effects of Sectoral and Economywide Policies" by Romeo M. Bautista and Clemen Gehlhar (January 1995)

No. 3- "International Trade, Regional Integration and Food Security in the Middle East" by Dean A. DeRosa (January 1995)

No. 4- "The Green Revolution in a Macroeconomic Perspective: The Philippine Case" by Romeo M. Bautista (May 1995)

No. 5- "Macro and Micro Effects of Subsidy Cuts: A Short-Run CGE Analysis for Egypt" by Hans Löfgren (May 1995)

No. 6- "On the Production Economics of Cattle" by Yair Mundlak, He Huang and Edgardo Favaro (May 1995)

No. 7- "The Cost of Managing with Less: Cutting Water Subsidies and Supplies in Egypt's Agriculture" by Hans Löfgren (July 1995, Revised April 1996)

No. 8- "The Impact of the Mexican Crisis on Trade, Agriculture and Migration" by Sherman Robinson, Mary Burfisher and Karen Thierfelder (September 1995)

No. 9- "The Trade-Wage Debate in a Model with Nontraded Goods: Making Room for Labor Economists in Trade Theory" by Sherman Robinson and Karen Thierfelder (Revised March 1996)

No. 10- "Macroeconomic Adjustment and Agricultural Performance in Southern Africa: A Quantitative Overview" by Romeo M. Bautista (February 1996)

No. 11- "Tiger or Turtle? Exploring Alternative Futures for Egypt to 2020" by Hans Löfgren, Sherman Robinson and David Nygaard (August 1996)

No. 12- "Water and Land in South Africa: Economywide Impacts of Reform - A Case Study for the Olifants River" by Natasha Mukherjee (July 1996)

No. 13- "Agriculture and the New Industrial Revolution in Asia" by Romeo M. Bautista and Dean A. DeRosa (September 1996)

No. 14- "Income and Equity Effects of Crop Productivity Growth Under Alternative Foreign Trade Regimes: A CGE Analysis for the Philippines" by Romeo M. Bautista and Sherman Robinson (September 1996)

No. 15- "Southern Africa: Economic Structure, Trade, and Regional Integration" by Natasha Mukherjee and Sherman Robinson (October 1996)

No. 16- "The 1990's Global Grain Situation and its Impact on the Food Security of Selected Developing Countries" by Mark Friedberg and Marcelle Thomas (February 1997)

No. 17- "Rural Development in Morocco: Alternative Scenarios to the Year 2000" by Hans Löfgren, Rachid Doukkali, Hassan Serghini and Sherman Robinson (February 1997)

No. 18- "Evaluating the Effects of Domestic Policies and External Factors on the Price Competitiveness of Indonesian Crops: Cassava, Soybean, Corn, and Sugarcane" by Romeo M. Bautista, Nu Nu San, Dewa Swastika, Sjaiful Bachri and Hermanto (June 1997)

No. 19- "Rice Price Policies in Indonesia: A Computable General Equilibrium (CGE) Analysis" by Sherman Robinson, Moataz El-Said, Nu Nu San, Achmad Suryana, Hermanto, Dewa Swastika and Sjaiful Bahri (June 1997)

No. 20- "The Mixed-Complementarity Approach to Specifying Agricultural Supply in Computable General Equilibrium Models" by Hans Löfgren and Sherman Robinson (August 1997)

No. 21- "Estimating a Social Accounting Matrix Using Entropy Difference Methods" by Sherman Robinson and Moataz-El-Said (September 1997)

No. 22- "Income Effects of Alternative Trade Policy Adjustments on Philippine Rural Households: A General Equilibrium Analysis" by Romeo M. Bautista and Marcelle Thomas (October 1997)

No. 23- "South American Wheat Markets and MERCOSUR" by Eugenio Díaz-Bonilla (November 1997)

No. 24- "Changes in Latin American Agricultural Markets" by Lucio Reca and Eugenio Díaz-Bonilla (November 1997)

No. 25*- "Policy Bias and Agriculture: Partial and General Equilibrium Measures" by Romeo M. Bautista, Sherman Robinson, Finn Tarp and Peter Wobst (May 1998)

No. 26- "Estimating Income Mobility in Colombia Using Maximum Entropy Econometrics" by Samuel Morley, Sherman Robinson and Rebecca Harris (Revised February 1999)

No. 27- "Rice Policy, Trade, and Exchange Rate Changes in Indonesia: A General Equilibrium Analysis" by Sherman Robinson, Moataz El-Said and Nu Nu San (June 1998)

No. 28*- "Social Accounting Matrices for Mozambique - 1994 and 1995" by Channing Arndt, Antonio Cruz, Henning Tarp Jensen, Sherman Robinson and Finn Tarp (July 1998)

No. 29*- "Agriculture and Macroeconomic Reforms in Zimbabwe: A Political-Economy Perspective" by Kay Muir-Leresche (August 1998)

No. 30*- "A 1992 Social Accounting Matrix (SAM) for Tanzania" by Peter Wobst (August 1998)

No. 31*- "Agricultural Growth Linkages in Zimbabwe: Income and Equity Effects" by Romeo M. Bautista and Marcelle Thomas (September 1998)

No. 32*- "Does Trade Liberalization Enhance Income Growth and Equity in Zimbabwe? The Role of Complementary Polices" by Romeo M. Bautista, Hans Lofgren and Marcelle Thomas (September 1998)

No. 33- "Estimating a Social Accounting Matrix Using Cross Entropy Methods" by Sherman Robinson, Andrea Cattaneo and Moataz El-Said (October 1998)

No. 34- "Trade Liberalization and Regional Integration: The Search for Large Numbers" by Sherman Robinson and Karen Thierfelder (January 1999)

No. 35- "Spatial Networks in Multi-Region Computable General Equilibrium Models" by Hans Löfgren and Sherman Robinson (January 1999)

No. 36*- "A 1991 Social Accounting Matrix (SAM) for Zimbabwe" by Romeo M. Bautista and Marcelle Thomas (January 1999)

No. 37- "To Trade or not to Trade: Non-Separable Farm Household Models in Partial and General Equilibrium" by Hans Löfgren and Sherman Robinson (January 1999)

No. 38- "Trade Reform and the Poor in Morocco: A Rural-Urban General Equilibrium Analysis of Reduced Protection" by Hans Löfgren (January 1999)

No. 39- " A Note on Taxes, Prices, Wages, and Welfare in General Equilibrium Models" by Sherman Robinson and Karen Thierfelder (January 1999)

No. 40*- "Parameter Estimation for a Computable General Equilibrium Model: A Maximum Entropy Approach" by Channing Arndt, Sherman Robinson and Finn Tarp (February 1999)

No. 41- "Trade Liberalization and Complementary Domestic Policies: A Rural-Urban General Equilibrium Analysis of Morocco" by Hans Löfgren, Moataz El-Said and Sherman Robinson (April 1999)

No. 42- "Alternative Industrial Development Paths for Indonesia: SAM and CGE Analysis" by Romeo M. Bautista, Sherman Robinson and Moataz El-Said (May 1999)

No. 43*- "Marketing Margins and Agricultural Technology in Mozambique" by Channing Arndt, Henning Tarp Jensen, Sherman Robinson and Finn Tarp (July 1999)

No. 44- "The Distributional Impact of Macroeconomic Shocks in Mexico: Threshold Effects in a Multi-Region CGE Model" by Rebecca Lee Harris (July 1999)

No. 45- "Economic Growth and Poverty Reduction in Indochina: Lessons From East Asia" by Romeo M. Bautista (September 1999)

No. 46*- "After the Negotiations: Assessing the Impact of Free Trade Agreements in Southern Africa" by Jeffrey D. Lewis, Sherman Robinson, and Karen Thierfelder (September 1999)

No. 47*- "Impediments to Agricultural Growth in Zambia" by Rainer Wichern, Ulrich Hausner, and Dennis K. Chiwele (September 1999)

No. 48- "A General Equilibrium Analysis of Alternative Scenarios for Food Subsidy Reform in Egypt" by Hans Löfgren and Moataz El-Said (September 1999)

No 49*- "A 1995 Social Accounting Matrix for Zambia" by Ulrich Hausner (September 1999)

No. 50- “Reconciling Household Surveys and National Accounts Data Using a Cross Entropy Estimation Method” by Anne-Sophie Robilliard, and Sherman Robinson (November 1999)

*TMD Discussion Papers marked with an "*" are MERRISA-related papers.

Copies can be obtained by calling, Maria Cohan at 202-862-5627 or e-mail m.cohan@cgiar.org