TMD DISCUSSION PAPER NO. 97

ESTIMATION OF A REGIONALIZED MEXICAN SOCIAL ACCOUNTING MATRIX: USING ENTROPY TECHNIQUES TO RECONCILE DISPARATE DATA SOURCES

Rebecca Lee Harris

Globalization Research Center

University of South Florida

Trade and Macroeconomics Division

International Food Policy Research Institute

2033 K Street, N.W.

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

September 2002

Prepared for IO Workshop in Guadalajara, Mexico, September 2-4, 1999. Rebecca Lee Harris, Globalization Research Center, University of South Florida

Abstract

This paper presents the construction of a 1996 regionalized Social Accounting Matrix (SAM) for Mexico. The SAM differentiates production across five regions, four rural and a fifth "national" urban region. The rural regions are differentiated by their agricultural production technologies. There are three households in each region, disaggregated by income level, so that the SAM can be used in studies of income distribution. The data come from a variety of sources, including Mexico's System of National Accounts, the National Survey on Household Income and Consumption, and the Secretary of Agriculture, Livestock and Rural Development. As a result, the data are not consistent and the "adding up" constraints of the SAM are not met. The SAM is then estimating using entropy techniques to incorporate the data in a consistent way.

Table of Contents

I. Introduction

This paper describes the construction of the social accounting matrix for the Mexican economy in 1996. This SAM will be used as the underlying data framework for a regionalized computable general equilibrium (CGE) model. The SAM regionally disaggregates production, factor markets and households, so that it is able to capture differences among 5 regions in a "top-down" approach: rather than having complete regional SAMs, the model only regionally disaggregates production and factor markets and households. While ideally the SAM would have more information on marketed surplus per region, this information is not available.

Each rural region has 6 agricultural production activities which are destined to national commodity markets, plus one national livestock-forestry-fisheries sector. The urban region contains 14 other production sectors. Each rural region has its own agricultural labor and its own irrigated and non-irrigated land. The urban region employs 4 labor types: professional, white collar, blue collar and unskilled/informal. One capital factor is used by all regions. Within each region, there are three households, defined as Poor, Medium and Rich, for a total of 15 households. The SAM also contains an account for enterprises, government, savings-investment, and the world.

The SAM accounts for all income and expenditure transactions of all sectors and institutions in an economy.[1] It is a square matrix, in which the rows represent the receipts and the columns represent the expenditures. Thus, since total expenditures must equal total income for each agent, the row sums must equal their corresponding column sums. Income or expenditure may be described as:

[1] For a detailed discussion on SAMs, see Pyatt and Round (1985).

$$y_i = \sum_j T_{i,j} = \sum_j T_{j,i}$$

where y_i is total income or expenditure of account *i* and $T_{i,j}$ is the cell entry representing the payment from column account *j* to row account *i*. The $T_{i,j}$ make up the transaction matrix, T. The SAM coefficient matrix, A, is composed of the SAM coefficients, $A_{i,j}$, or

$$A_{i,j} = \frac{T_{i,j}}{y_j}$$

each cell entry divided by its respective column sum:

The column sums of A equal one by construction. In matrix notation, the model is

$$y = A\,x$$

written as

in which $y = x$.

Assembling the SAM consists of gathering data from a variety of sources, including input-output (IO) data, national accounts data and household survey data. These data are often inconsistent with each other, so that it is unlikely that the raw data will lead to a balanced SAM without further manipulation.

Building a SAM consists of the following steps: first a "macro SAM" is built, which comes mainly from national accounts data and defines the macro flows among the major players of the economy. Then the "micro SAM" is built, which can be thought of as a finer disaggregation of the macro SAM. Because the micro SAM's entries come from disparate sources, it is most likely unbalanced and may even be inconsistent with the macro SAM. To some extent, practical knowledge as well as economic intuition can be used to fix some of these inconsistencies. This first step of assembling the micro

SAM results in what is known as the “proto-SAM,” which is probably still unbalanced, but as consistent as possible using the available data. When the proto-SAM is satisfactorily adjusted, there are two approaches which can be taken to properly balance it. The traditional RAS approach is appropriate only in the case of starting with an older, but consistent SAM, in which the new column totals are known. The more appropriate method in the current problem is to use entropy estimation techniques, which can take full advantage of all of the known information (i.e. particular flows within the SAM) while accounting for measurement errors, lack of data, and other inconsistencies.

The organization of this paper follows these steps of constructing the SAM: Section 2 describes the assembly of the macro SAM, which is disaggregated into the proto-SAM in Section 3. Section 4 discusses the theory of entropy and its application to balancing the proto-SAM. Section 5 makes some concluding comments.

II. The Macro SAM

The macro SAM accounts for the major players in the economy. Typically these are producers, factors of production, institutions, and a foreign component. Table 1 shows the framework of the macro SAM that is used in the Mexico model and Table 2a shows the actual values for the Mexican macro SAM. Production is split into two types of categories: "activities" and "commodities." The activity account may be thought of as the domestic producers account. On the column, it consists of intermediate inputs, value added and value added and producer taxes. Along the row, it accounts for domestic production and home consumption.[2] Because there is no home consumption in the macro accounts, this is calculated in the construction of the micro SAM (described

[2] Some SAMs place exports in the activity row - i.e. the rest of the world is purchasing from the activity account. In that case, the commodity purchases from the activity account would represent domestic sales (marketed output minus sales abroad). In the current model, as will be described below, exports will be sold from the commodity account. Thus, commodity purchases from the activity account represent all of marketed (ie. net of home consumption) domestic production.

below) and then subtracted from the purchased private consumption entry in the macro SAM. The sum of the activity purchases or income is production at factor costs, or gross domestic output, which in the Mexico SAM equals 4,167 billion pesos.

Table 1. Macro SAM

	Expenditures										
Receipts	**Activities**	**Commodities**	**Labor**	**Capital**	**Land**	**Households**	**Enterprise**	**Government**	**Savings-Investment**	**World**	**Total**
Activities		Domestic Production				Home Consumption					Production
Commodities	Intermediate Inputs					Private Consumption		Government Consumption	Investment	Exports	Total Demand
Labor	Value-added (wages)									Factor Income from Abroad	Gross National Product at Factor Prices
Capital	Value-added (rent)										
Land	Value-added (land returns)										
Households			Wages and Salaries		Land Returns		Operating Surplus	Transfers		Transfers from Abroad	Household Income
Enterprise				Profits				Transfers		Foreign Capital	Enterprise Income
Government	Value-added Taxes and Producer Taxes	Sales Taxes, Tariffs, and Export Taxes	Soc.Sec. Taxes			Income Taxes	Corporate Taxes			Foreign Borrowing	Govern-ment Income
Savings-Investment						Household Savings	Corporate Savings	Government Savings		Capital Transfers	Total Savings
World		Imports		Capital Transfers Abroad				Foreign Interest Payments			Imports
Total	Production at Factor Cost	Total Absorption	Labor Outlays	Capital Outlays	Land Outlays	Household Expenditures	Enterprise Expenditure	Government Expenditure	Total Investment	For. Exchange Inflow	

Table 2a. Initial Macro SAM for Mexico, 1996. Billions of Pesos.

Receipts	Activities	Commodities	Labor	Capital	Land	Households	Enterprise	Government	Savings-Investment	World	Total
Activities		4167									4167
Commodities	1799					1691		189	481	801	4961
Labor	729										729
Capital	1496										1496
Land	62										62
Households			683		62		1116	121		6	1988
Enterprise				1389				49			1438
Government	81	55	46			69	141				393
Savings-Investment						227	182	-11		84	481
World		739		107				45			891
Total	4167	4961	729	1496	62	1988	1438	393	481	891	

Table 2b. Data Sources For Macro SAM by Cell Entry[a]

INEGI, "Sistema de Cuentas Nacionales"

cell:

(Commodities, Activities)
(Labor, Activities)
(Capital, Activities)
(Land, Activities)
(Gov, Activities) - producer tax
(World, Commodities)
(Households, Labor)
(Gov, Labor)
(World, Capital)
(Commodities, Households)
(Commodities, World)
(Households, World)

Bank de Mexico

cell:

(Govt, Activities) - value added tax
(Govt, Commodities) - sales tax, export tax
(Commodities, Govt)
(Households, Govt)
(Enterprises, Govt)
(World, Govt)
(Savings-Investment, Govt)

Robinson, et al (1995)

cell:

(Govt, Households)
(Govt, Enterprises)

[a]Not all of the non-zero cells are cited either because they are calculated from previous cells or because they are used to balance the SAM. Land returns come from micro SAM data (see text).

The commodity account represents the market for goods. On the column, it shows where domestic agents buy the goods from: domestic production and imports (and import tariffs and other taxes). The column total represents total absorption (as opposed to domestic absorption, which would net out exports) and equals 4,222 billion pesos in Mexico. The row shows the domestic agents buying goods in the form of intermediate goods (purchased by the activities, and often referred to as the "use" matrix) and final goods (purchased by households, government, capital and the rest of the world).

The distinction between activities and commodities highlights two features of an economy. First, it allows more than one type of activity to produce a given commodity — for example, in the Mexico case, there will be agricultural activities from different regions producing the same product which will feed into one national commodity. Similarly, one activity may produce multiple commodities, for example, a dairy activity may produce both a cheese and a milk commodity (though this is not done in the Mexico model). In both cases, different production technologies can be highlighted. Second, the separation of activities and commodities allows the distinction between the sources of commodities (as domestically produced or as imports) as well as between the destinations of the commodities (as domestic sales or as exports).

The factors' entry in the macro SAM has three columns and respective rows, for aggregated labor, capital, and aggregated land. In the case of Mexico, there is no aggregate data for returns to land, but rather returns to land are included in the returns to capital. This is amended by using estimates from the Global Trade Analysis Project (GTAP) based at Purdue University for the aggregate labor, capital and land ratios. Along the row, the factors receive value added payments from the activities account and factor income from abroad, and down the column, factors make payments to the institutions. Institutions typically are comprised of an aggregate household, the government, an enterprise account and the rest of the world. The labor account pays wages and salaries to the household and social security taxes to the government. The capital account pays some of its receipts to the enterprise account, which serves as the intermediary between the capital account and the households, and it sends some capital returns to the rest of the world. The land account pays its returns directly to the household.

The household receives factor income (though capital payments come indirectly from the enterprise account), government transfers and transfers from the rest of the world along the row and pays for consumption (including home consumption out of the activities account), taxes and savings down the column. The government receives the activities' taxes, which are the producer tax and the value added tax, the commodities' taxes — which include import tariffs, export taxes, and sales taxes — social security taxes from the labor account, and income taxes from households and enterprises. Any foreign borrowing is entered in the row as a payment from the rest of the world. The government purchases commodities, gives transfers to households and enterprises, and makes interest payments to the rest of the world. The government also saves or dis-saves, by its positive or negative payments to the capital account. The enterprise account collects value added capital, government transfers, and foreign capital along the row and distributes these revenues to households, the government, and the capital account down the column. The savings-investment account receives private savings, enterprise savings, government savings (or dis-savings), and foreign savings along the row and invests in the commodities account down the column. Finally, the rest of world row account receives payments for imports from the commodities account, and receives capital outflows and foreign payments from the government. Export receipts are paid to commodities in the column entry, along with foreign factor income, transfers and government borrowing.

The macro SAM is relatively easy to manage and balance. Most of the data comes from the same source (in the Mexican case, from the National Accounts data - see Table 2b) and because it is a small matrix (10x10), any discrepancies can be adjusted manually. The micro SAM is a different story.

III. The Micro SAM

The micro SAM, as its name suggests, is a much finer disaggregation of the macro SAM. The data comes from a variety of sources and as a result, often does not equal the macro totals. However, since the macro data is usually believed to be most accurate and consistent, the micro data is often scaled so that in the aggregate it equals the appropriate macro total (often referred to as the "control" total).

Initially, the SAM was constructed following the Mexican National Accounts and had 96 production sectors in the economy, meaning that there were 96 activities and 96 commodities. However, this makes for an unwieldy SAM, and furthermore, not all of the sectors need to be disaggregated. For the current study, there are 21 national sectors (see Table 3 for a listing of these sectors). The model will have a heavy emphasis on agriculture, so 7 accounts are from agriculture and livestock production, and 6 accounts from the processed foods sector. Accounts are disaggregated according to their importance in output and in trade and vertical integration, so, for example, *Maize* is a separate account from *Maize Manufacturing*. Additionally, some products are isolated due to their importance in the policies on which later CGE analysis will focus, such as maize or beans. For each of the agricultural crops, there are four activities for each sector, since agricultural activities are regionalized into four rural regions. Thus, the *Maize* sector is associated with activities called *Maize-North, Maize-Central, Maize-Southeast,* and *Maize-Southwest*. While production of the livestock-forestry-fishery composite comes from all regions, its technology is not differentiated according to region (due to data unavailability) so there is only one activity for this sector. There are thus 24 regionalized agricultural activities and 39 activities altogether. The commodity accounts follow the sectoral listing exactly, for a total of 21 commodities.

Table 3. National Sectors in Model[a]

Table 3. National Sectors in Model a

1. Maize
2. Wheat
3. Beans
4. Other Grains (Sorghum, Barley)
5. Fruits and Vegetables
6. Other Crops (Tobacco, Hemp, Cotton, Cocoa, Sugar, Coffee, Soy, Safflower, Sesame and Others)
7. Livestock/Forestry/Fisheries (Bovines, Goats, Sheep, Bees, Poultry and Others, Forestry and Fisheries)
8. Dairy
9. Prepared Fruits and Vegetables
10. Wheat Manufacturing
11. Corn Manufacturing
12. Sugar Manufacturing
13. Other Processed Foods (Coffee Manufacturing, Processed Meats, Oils and Fats, Feeds, Alcohol, Beverages and Others)
14. Light Manufacturing (Lumber, Wood, Paper, Print, and Cigar Manufacturing, Soft Fiber Textiles, Hard Fiber Textiles, Other Textiles, Leather, Apparel)
15. Intermediates (Chemicals, Synthetics, Rubber, Glass, Cement, Fertilizers, Other Chemicals, Oil Refining, Oil and Gasoline, Petrochemicals, Coal, Iron, Non-Ferrous Metal, Sand/Gravel, Minerals)
16. Consumer Items (Pharmaceuticals, Soaps, Plastic, Metal Furnishings, Household Appliances, Electronic Equipment, Automobiles and Parts)
17. Capital Goods (Metal Products, Metal Manufacturing, Non-Electronic Machines, Electronic Machines, Other Electric Goods, Transportation Materials, Mineral Manufacturing, Iron Manufacturing, Non-Ferrous Metal Manufacturing, Others)
18. Professional Services (Professional Services, Education, Medical, Finance/Real Estate, Public Administration and Defense, Electricity, Gas and Water)
19. Other Services (Other Services, Restaurants
20. Construction
21. Commerce, Trade and Transportation

[a] Note that there are four activities for each of the agricultural crop sectors (sectors 1- 6): one for each region. Otherwise, the activities are the same as these sectors. The commodities are the same as these sectors.

It should be noted that this disaggregation can be changed to accommodate a different level of analysis. For example, if this SAM were used to explore industrial policies, the agricultural sectors would be more highly aggregated and the urban sectors would have greater detail. This SAM can also easily be aggregated up into a purely national SAM, with only one activity per commodity. Finally, it would not be difficult to make the SAM more finely disaggregated, up to the 96 sectors from the National Accounts data, though it becomes more computationally costly as the number of sectors increases.

Next, a more detailed description of the sectors of the micro SAM is presented, in the same order as the macro accounts.

A. Activity Account

For each sector, the first focus is on the activity column, in order to determine gross output per sector (the column sum). This is done using a combination of the input-output (IO) table from 1985 and using national accounts data from 1996. For all of the non-agriculture sectors and the livestock-forestry-fishing sector, there is 1996 data for intermediate demand, value added and indirect taxes, the three components which comprise gross output.

For the agriculture and livestock sectors, there is 1996 data of these three components for *aggregate* agriculture and *aggregate* livestock. However, for each *individual* sector, there is only data on gross output and land returns. Intermediate demand, value added payments to labor and capital, and indirect taxes are calculated by applying the IO coefficients from 1985 (ie. the value of the item in question divided by gross output, all for 1985) to gross output for 1996. Since there is no land share in the IO table, the capital payments in the 1985 data are assumed to include land returns, so the land returns (described below) from the 1996 SAGAR data are netted out of the calculated capital returns. Because the IO shares from 1985 reflect a different technology from more recent years, the summation of this calculated data across sectors leads to a different allocation of payments to intermediate goods, value added and indirect taxes than what is reported in the 1996 data for aggregate agriculture and aggregate livestock. This discrepancy will be adjusted in the entropy operation.[3] After the total intermediate demand by activity is calculated, that demand is allocated across the commodity sectors. Here, the IO coefficients are applied to the total intermediate demand by sector, to get the flows between activity and commodity.

The land return data is calculated by taking crop data on gross output and land use from SAGAR. First, relative returns between irrigated and non-irrigated land per crop are calculated by assuming that the relative return to dry land is equal to the number of hectares of dry land utilized, and that the relative return to irrigated land is equal to the number of hectares of irrigated land utilized, multiplied by the ratio of dry land yield to irrigated land yield. These data are then scaled so that the sum for irrigated and non-irrigated land equal the aggregate total return to land.

There are four "rural" regions in the model (see Table 4) as well as one "urban" region which represents urban areas anywhere in the country. The regions are put into the SAM so that the CGE analysis can look at how national or regional policies will affect different parts of the country. Because the CGE model will be used to analyze agricultural policies, the country should be divided into sensible agro-economic regions.

[3] While the entropy procedure will reconcile many such inconsistencies, some manual adjustment is necessary in this part of the SAM's construction. In particular, the value-added shares given by the national accounts data make Mexico look like a very capital intensive country. This is a common problem in data coming from developing countries, in which it is very difficult to measure labor payments - due to informal markets, under-reporting, self employment, etc. Bearing this in mind, the best solution is to cut the capital returns in half in the overly capital intensive sectors (here, the food processing, light manufacturing and service sectors) and redistribute that share to the labor returns. A similar issue occurs in some of the agricultural sectors, in which the *Fruits and Vegetables* and *Other Crops* sectors appear too capital intensive and so some capital is reallocated to land, following GTAP figures.

Mexico is a diverse country in this regard, making it difficult to keep a manageable number of regions. These four regions would undoubtedly not satisfy an agronomist, but each one has enough specific characteristics to keep the model interesting without being overly cumbersome. Generally, the northern part of Mexico has higher value crops and higher incomes. The southern part of the country tends to be poorer, with more subsistence farming and less commercial agriculture. The North region is comprised of states which tend to produce agriculture for export, and the crops are high-value. Popular crops include fruits and vegetables, and wheat. The Central region is a combination of arid and temperate zones, with corn and beans the major crops. There is less subsistence farming, more use of irrigated land, and the climate is dry. This area is a combination of irrigated and non-irrigated land. The Southwest region is subhumid tropical, along the Pacific coast. Corn is a major crop in these states and there is a much lower usage of irrigation. Finally, the Southeast region contains the poorest states, the climate is tropical and corn, rice and coffee are the important crops.

The agricultural activities are divided into regions, so that there are four regionalized agricultural activities for each of the formerly national ones. The land payments are divided up by crop and region according to state-level data aggregated to the regions.[4] Some estimates from INIFAP (Instituto Nacional de Investigaciones Forestales Agricolas y Pecuarias) are used to get land-labor ratios, which are then applied to divide up the labor payments by region. While there is data for at least one state in each region, these data are somewhat sparse and not necessarily representative of the entire region. Nevertheless, it provides an estimate for differentiating value added by region. There is no data available for intermediate demand by regionalized crop, so these figures are split according to total output per crop. Similarly, the producer tax (which is actually a subsidy for most agricultural products) is allocated according to total output per crop, while the value-added tax is distributed according to total value added per crop.

[4] This data comes from SAGAR (1996).

Finally, capital is determined as the residual of total output minus the rest of value added, the intermediate demands and the taxes.

Table 4. Rural Regions

1. North
   - -Baja California Norte
   - -Baja California Sur
   - -Sonora
   - -Sinaloa
   - -Chihuahua
   - -Coahuila
   - -Nuevo Leon

2. Central
   - -Durango
   - -Zacatecas
   - -Aguascalientes
   - -San Luis Potosi
   - -Guanajuato
   - -Queretaro
   - -Hidalgo
   - -Tlaxcala
   - -Puebla
   - -Tamaulipas

3. Southwest
   - -Nayarit
   - -Jalisco
   - -Colima
   - -Michoacan
   - -Estado de Mexico
   - -Distrito Federal
   - -Guerrero
   - -Morelos

4. Southeast
   - -Veracruz
   - -Oaxaca
   - -Chiapas
   - -Tabasco
   - -Campeche
   - -Yucatan
   - -Quintana Roo

In the regionalized SAM, irrigated and non-irrigated land are both classified into the four regions, for a total of eight land-types. Each crop can only pay to the land that is in its own region. Likewise, labor payments are divided by region, such that each crop will only pay labor in its own region. It is assumed that labor payments from agricultural activities are only destined to agricultural labor.

The activity column totals sum to gross output by sector, which is then used to construct domestic production by sector in the activity row. Home consumption, which comes out of the activity row, reflecting that it is "purchased" at producer prices, is netted out of gross output by sector and the difference is placed in the domestic production cell. Thus, by construction, the activity row and column will balance.

B. Commodity Accounts

Moving to the commodity row, private consumption, government consumption and investment by sector are all determined by applying the 1985 IO shares to the 1996 macro data. The export data comes from 1996 data from GTAP and must be scaled so that the sum equals the aggregate export total for 1996. Along the row, import tariffs and imports by sector are determined with the 1996 trade data from GTAP. The commodity accounts will be unbalanced, in part because of the forced balancing of the activities accounts (in which the difference between each activity's column and row total is pushed into the commodity account via the domestic production cell). In addition, the IO table is obviously outdated, and will not reflect all changes in technology, or changes in the ratio of intermediate demand to final demand of a particular product. Hence, part of the manual balancing of the commodity columns and rows will consist of judging how these changes would affect the IO table. For example, the consumer durables commodity account will be unbalanced, largely because the IO table does not reflect the increased private final demand for these products. In the cases in which intermediate demands — ie., a payment from activities to commodities — have to be adjusted, this requires rebalancing the activity column as well.

After the commodities and activities transactions are more-or-less complete (since the micro SAM will still be unbalanced), the other actors from the macro SAM can be broken down. In particular, aggregate labor will be divided by skill level and households will be divided by region and income level.

C. Labor Categories

There will be 8 labor categories: professional, white collar, blue collar, unskilled/informal (referred to as "unskilled" in the model), and 4 agricultural types — one for each rural region — as further described in Table 5. The micro SAM is used to map the income that the labor categories receive from the production sectors and then direct it to the different households. The mapping is determined using data from the consumption and income survey, ENIGH (Encuesta Nacional de Ingresos y Gastos Hogares) by the National Statistics Institute, INEGI (Instituto Nacional de Estadística Geografía e Informática), in which each household reports all of its income according to both skill category and production sector.

Table 5. Labor Categories

1. Professional

   -Professionals

   -Technicians

   -Teachers and Educators

   -Artists, Entertainers, Professional Athletes

   -Armed Forces, Protective Service Workers

2.Agricultural (1 per rural region)

3. Blue Collar

   -Factory Workers

   -Machine Operators

   -Transportation and Machine Drivers

4. White Collar

   -Factory Supervisors and Inspectors

   -Department heads, Coordinators, Supervisors of Administration and Services

   -Administrative Workers

-Workers in Sales and Commerce

5. Unskilled/Informal

-Unskilled

-Informal

-Street Vendors

-Domestic Servants

Thus, each production sector's payments to aggregate labor get divided among the labor categories according to the shares determined from the income survey.[5] Note that agricultural activities give labor payments only to agricultural laborers (of the same region), and non-agricultural activities do not make any labor payments to agricultural laborers.

D. Households

The households are distinguished by location and income statues. There are 5 regions and 3 income levels, for a total of 15 households. For this study, the definition of "rural" differs from the commonly used definition, in which a rural region is one with fewer than 2,500 people.

Under this traditional definition, 75 percent of the Mexican population is urban and the urban region earns 90 percent of national income. For the purposes of this study, that means that the 12 rural households (3 households for each of 4 rural regions) must split up the remaining 10 percent of income. This implies that some households would earn such small amounts of income as to look insignificant. The rich rural households would effectively earn zero income and the other rural households would earn less than percent of income each. Given that the focus of the CGE analysis is on agriculture and rural areas, it seems to be logical to expand the definition of "rural."

[5] Due to data unavailability, the activities' payments to labor categories come from the household survey, though an industry survey would be more accurate.

This is achieved by changing the definition of "rural" to include households living in towns with up to 15,000 people. This seems reasonable for several reasons. First, in Mexico, a town of 15,000 people, and potentially fewer than 3,000 families, still has a large amount of agricultural activities. Secondly, by Mexico standards, 15,000 is not a "big" town. The metropolitan region Mexico City has 15 *million* people.

Under this new definition, about 61 percent of the population is urban, earning 81 percent of national income. Since home consumption is considered to be higher in rural regions, it is important to see how this indicator would change under the new definition. Among households living in areas of less than 2,500 people (the traditional definition of rural), almost 8 percent of total current expenditures is from home consumption. Among those living in areas with 2,500 to 15,000 people, home consumption makes up about

3 percent of total spending. If these two categories are combined, the weighted home consumption averages at percent of total spending. While this means that home consumption is lower under this new definition, contrast it to those living in areas of more than 15,000 people: here, self-consumption is less than percent of total spending. Thus, from a self-consumption point of view, it would appear that some data would be improperly aggregated if these small towns were merged with larger towns and cities.

Household income levels are split into just three categories: rich, medium or poor. The delineation among the categories comes from national data, such that the poor are those in the lowest 40 percent income bracket of the entire country, regardless of their location, the medium earn the next 40 percent of income and the rich households earn the top 20 percent of income. Thus, there are 15 different households in the model; 3 urban and 12 rural. Thus welfare effects in the CGE model will be captured not just among income categories but also among regions. Labor payments by category get distributed to the different households (known as the "allocation matrix") according to the household survey data. Obviously, households can only receive the agricultural labor payments from the region in which they are located.

Land payments to households are also region specific. There is no data available for the distribution of land payments, so here, the assumptions from Levy and van Wijnbergen (1992) are "liberally" adapted.[6] In their CGE model of Mexico, they distribute irrigated and non irrigated land to different farmer types, not income types, in the following manner: All irrigated land returns go to "irrigated farmers." Half of the dryland farmers are subsistence farmers, and they get 45 percent of dryland returns. The other half of dryland farmers get 55 percent of dryland returns. In the current study, this is adapted by assuming that Rich farmers get all of irrigated returns and Medium farmers receive 55 percent of dryland returns and Poor farmers, roughly equivalent to subsistence farmers, get 45 percent of dryland returns. Note that the distribution is the same across regions, since the Levy and van Wijnbergen model is a national one, though of course the magnitudes of returns are different, as are the ratios to other factors.

[6] Ideally, there would be data to allow rural households to be characterized by land ownership or another agriculture-related measure. Indeed, even the figures from the Levy and van Wijnbergen study, are based on general data and strong assumptions.

Households receive capital payments via the enterprise account. This is calculated as the residual between household spending and all other household income (labor and land payments, plus government and foreign transfers). This calculation highlights another difficulty in constructing SAMs for countries with imperfect data. Since all of the income sources are not well specified in the data, initially the enterprise income was a huge source of income for many of the households. While enterprise income may be related to informal capital earnings or informal business earnings (as well as earnings from formal sector capital), in some instances more than 60 percent of a poor rural household's income would come from the enterprise account. In order to rectify this, factor payments from some activities were reallocated again, so that there would be more labor income available to the households — particularly unskilled labor. This is a reasonable conjecture, if most of the under-reported income comes from informal sources of labor.

The share of a household-type's consumption in the total consumption of a commodity is determined through the consumption data by ENIGH. The ENIGH data considers 565 categories of consumption goods which must be aggregated to fit the 22 sectors, except for the sectors which are not consumer final demand categories, (ie. intermediate goods, capital goods). Then, the consumption share is applied to the total consumption of the commodity, as given earlier from the IO data.

E. Other Institutions

The government's purchases of each commodity is determined according to the shares given by the 1985 IO Table, multiplied by the macro SAM entry of government purchases of the aggregate commodity. Government transfers to each household are calculated by applying the shares given by the ENIGH survey to the macro SAM total. The government payments to the rest of the world and to the savings-investment account are the same as in the macro SAM. As for government receipts, these are simply a collection of the various tax payments listed earlier.

The rest of the world purchases exports from the commodity accounts according to the proportions given by the GTAP database, applied to the macro SAM totals. Rest of the world transfer payments to households are derived from the ENIGH survey, and foreign savings is the same as in the macro SAM. Imports from the rest of the world come from the GTAP database, while the other rest of the world receipts (from the capital factor and the government) equal the macro SAM totals. The savings-investment account shows investment in each commodity down the column, calculated using the IO coefficients. Receipts, from government and the rest of the world, are from the macro SAM totals.

When all of these data are entered, the proto-SAM is unbalanced, but it is as consistent as possible given the data. In the next section, the methods for balancing the proto-SAM are discussed.

IV. Entropy Approach to Estimating Micro SAM

As can be seen from the above description, the entries of the micro SAM come from a variety of sources, ranging from an outdated input-output table, to 1994 census data, to current national accounts data. The agricultural data comes from a current publication of the Ministry of Agriculture, but it has discrepancies with the national accounts data. Thus, it comes as no surprise that the micro SAM which results from these disparate sources is not balanced.

The traditional solution for balancing a SAM (or any matrix with known row and column sums) is to use the RAS approach. The RAS approach is typically used in updating a SAM, in which the new row and column sums, y*, are known. The RAS procedure finds a new coefficient matrix, A^*, based on the original coefficient matrix, $\overline{A}$, which produces a new transactions matrix, T^*, which is consistent with the new row and column sums. Essentially, the RAS procedure iteratively adjusts the row and column entries proportionately until the totals are reached.

The RAS approach has several drawbacks. It assumes that the initial SAM upon which it is based is balanced and that there is no measurement error in the new row and column sums. Also, there is no way to incorporate other knowledge besides the row and column sums. For example, in addition to knowing the total production of a particular sector, there may be specific information on payments to the components of value added for that sector. These deficiencies are rectified with the entropy econometrics estimation procedure as discussed by Golan, Judge and Miller (1996), which has been applied to SAM estimation in Robinson, Cattaneo, and El-Said (1998).[7]

A. Entropy Theory

The entropy technique, which originally comes from information theory, is a way of solving underdetermined estimation problems,[8] using as much prior information as possible. It is motivated from the ill-posed inverse problem

$$\mathbf{y} = X\mathbf{p}$$

in which **y** is a T-dimensional vector of observations, **p** is an *n*-dimensional vector of unknowns and X is a known T x *n* matrix. If *n* is larger than T, **p** cannot be solved using traditional econometric techniques without very strict assumptions.

---

[7] McDougall (1999) demonstrates that the RAS approach is, in fact, an entropy theoretic model and suggests that it is a superior method to the entropy framework which will be shown here. However, RAS is only an appropriate tool when row and column sums are known (and believable) and when there is no further information on flows within the matrix. The RAS methodology cannot incorporate new flows information and still maintain the proportional changes in both the column and row coefficients (i.e., the bi-proportionality condition).

[8] Note that in the case of SAM estimation, the *n*-by-*n* cells of the matrix must be estimated with only 2*n* -1 independent row and column adding up restrictions.

The information theoretic problem that was posed by Shannon (1948) was to find a measure for the "uncertainty" of a probability distribution **p**. Suppose that the outcome of an event, for example, the rolling of a die, occurs with probability p. If there are *n* possible outcomes for the event, their associated probabilities — the probability distribution — are $p_1, p_2..p_n$, where all of the p_i are between 0 and 1, and the sum of all of the p_i equals 1. Different probability distributions have different levels of uncertainty associated with them. For example, in the die-rolling example, if the probability distribution for the face values {1,2,3,4,5,6} is {1/6,1/6,1/6,1/6,1/6,1/6,} — each face value occurs with equal probability — there is much less certainty than if it is {5/6,1/6,0,0,0,0}. In the latter case, it is fairly certain that a value of 1 will come up and this distribution gives very little information.[9] In the former case, uncertainty is very high: all numbers are equally likely to show. Indeed, as the probability distribution moves toward the uniform distribution, uncertainty is increased.

The entropy measure, defined by Shannon (1948) and Jaynes (1957), captures the "uncertainty" of a probability distribution, **p**, as follows:

$$H(p) = \sum_{i}^{n} p_i \ln \frac{1}{p_i} = -\sum_{i}^{n} p_i \ln p_i$$

with the definition that $p_i \ln p_i = 0$ if $p_i = 0$.

[9] This distribution has a low information content. That is, if the 1 does turn up, since it is expected, no new information is gained from this event occurring.

This measure has several desirable properties consistent with the concept of information content:[10] (1) $H(p)$ should not change if the p_i are reordered among themselves. (2) $H(p)$ is a continuous function, so that if the value of some p_i change by a

[10] See Kapur and Kesavan (1992) for further details.

small amount, the entropy will also change by a small amount. (3) If an impossible event (ie., $p_i = 0$) is included, H(p) should not change. (4) As long as p_i lies between 0 and 1, H(p) will have a non-negative value. (5) Any distribution in which one p_i equals 1 and the rest of the p_i equal 0 will have an entropy value of 0. This makes sense, since in this case, the outcome is known with complete certainty, and so uncertainty equals 0. For all other distributions, in which all of the probabilities are greater than zero, entropy is positive, indicating some level of uncertainty. (6) H(p) is a concave function, so its local maximum is a global maximum. Furthermore, if H(p) is maximized subject to the constraint that the p_i sum to one, the solution gives the uniform distribution, which, as noted above, is maximum uncertainty.

As more information is added to the entropy problem, uncertainty is reduced. In the die-rolling problem, if no other information is given, then the uniform distribution is the most uncertain one of all of the possible distributions. However, if more information is given, for example the mean value of the tosses is given, then the uncertainty is reduced, as the set of potential probability distributions shrinks.

The underlying philosophy in entropy estimation is that all of the information available should be utilized, but that information which is not certain or which is assumed without reasonable foundation should not be used.[11] Entropy, or uncertainty, should be maximized, given the constraints, because otherwise it would imply that some

[11] As a consequence, this approach, in contrast to standard statistical techniques, assumes very little about the error generating process and nothing about the functional form of the error distribution.

information which was not available was used in the estimation procedure. Thus the maximum entropy principle may be stated as: Out of all of the probability distributions which are consistent with the constraints, choose the one which maximizes the entropy metric.

The constraints in a maximum entropy problem are moment consistency constraints

$$\sum_{i=1}^{n} p_i \; f_t(x_i) = y_t \quad 1 \le t \le T$$

captured by:

which could include, for example, information on the mean or variance of the

$$\sum_{i=1}^{n} p_i = 1$$

distribution. The adding-up/normalization constraint is also included as:

The probability vector, **p**, is estimated by maximizing H(**p**), subject to these constraints,

$$L = -\sum_{i=1}^{n} p_i \ln p_i + \sum_{t=1}^{T} \lambda_t \left[y_t - \sum_{i=1}^{n} p_i \; f_t(x_i) \right] + \mu \left(1 - \sum_{i=1}^{n} p_i\right)$$

which is equivalent to maximizing the following Lagrangian:

where λ and μ are the Lagrange multipliers.[12] The first order conditions are as follows:

$$\frac{\partial L}{\partial p_i} = -\ln \hat{p}_i - 1 - \sum_{t=1}^{n} \hat{\lambda}_t \ f_t(x_i) - \hat{\mu} = 0$$

$$\frac{\partial L}{\partial \lambda_t} = y_t - \sum_{i=1} \hat{p}_i \ f_t(x_i) = 0$$

$$\frac{\partial L}{\partial \mu} = 1 - \sum_{i=1} \hat{p}_i = 0$$

The solution is:

[12] The Lagrange multipliers represent the marginal entropies; that is, they reflect how much entropy is reduced (increased) by relaxing (tightening) their respective constraints.

$$\hat{p}_i = \frac{\exp \sum_{t=1}^{T} \hat{\lambda}_t \ f \ (x_i)}{\Omega(\ \hat{\lambda}_1 \ \hat{\lambda}_2 \ldots \hat{\lambda}_t)}$$

$$\Omega(\hat{\lambda}) = \sum_{i=1}^{n} \exp\left[\sum_{t=1}^{\hat{T}lambda_t} f_t(x_i)\right]$$

where

Given the Lagrangian and its first order conditions, the Hessian to this problem is:

Since the Hessian is negative definite for $p_i > 0$, the solution is a unique global maximum. There is no closed-form solution for this formulation (since it depends only on the λs), so it can only be found numerically. However, from the Hessian, it is seen that the problem is poorly scaled, making it computationally very costly.

To ease this computational difficulty, the problem can be formulated in its dual. Starting from **y** = X**p**, the dual is constructed with the solution to the earlier Lagrangian as:

The Hessian of the dual objective **M**(λ) can be shown to be the variance-covariance matrix of the random variable **x** implied by $\hat{p}$. As such, the scaling is improved and it is much more computationally efficient to use than the primal of the maximum entropy problem.[13]

B. Cross Entropy

Maximum entropy is an appropriate tool when there is no prior information on the probability distribution, other than the moment consistency constraints and the adding up condition. Kullback and Leibler (1951) show that when there is a known prior of the distribution, q, the "cross entropy" solution is to minimize the entropy distance between the prior and the new distribution as:

$$I(p:q) = \sum_{i}^{n} p_i (\ln p_i - \ln q_i) = \sum_{i}^{n} p_i \ln (p_i / q_i)$$

which will be seen as the measure used in the SAM estimating problem. If the prior distribution is uniform — in other words, it contains no information — then minimizing cross entropy is equivalent to maximizing entropy. Using a prior distribution is consistent with the entropy philosophy stated above; believing that the prior distribution contains concrete information, the objective is to find a new set of probabilities which will

[13] See Golan and Judge (1996) for further details.

minimize the entropy distance between the new set and the old. Any other solution would contain unavailable information.

The Kullback-Leibler cross entropy distance in equation (13) is minimized, subject to the same moment consistency and adding-up constraints as in the earlier case, leading to

$$L = \sum_{i=1}^{n} p_i \ln(p_i / q_i) + \sum_{t=1}^{T} \lambda_t \left[y_t - \sum_{i=1}^{n} p_i \; f_t (x_i) \right] + \mu \left(1 - \sum_{i=1}^{n} p_i \right)$$

the Lagrangian:

$$\hat{p}_i = q_i \frac{\exp \sum_{t=1}^{T} \hat{\lambda}_t \; f \; (x_i)}{\Omega (\hat{\lambda}_1 \, \hat{\lambda}_2 \ldots \hat{\lambda}_t)}$$

The solution is:

$$\Omega (\hat{\lambda}) = \sum_{i=1}^{n} q_i \; \exp \left[\sum_{t=1}^{T} \hat{\lambda}_t \; f_t \; (x_i) \right]$$

where

Again, this formulation may be stated in the dual for computational efficiency.

C. Errors in Variables Approach

In either the maximum entropy or cross entropy technique, errors in measurement can

$$y = X \; p + e$$

incorporated by adding an error vector, **e**, to the problem:

The errors are a weighted average of known constants v:

$$e_i = \sum_w W_{i,w} \bar{v}_{i,w}$$

subject to the constraint that the weights, W, sum up to one. *w* is the set of weights, which determine the distribution of the error. The constants, v, define the support set for the errors. The weights can be thought of as probabilities to be estimated. Focusing on the cross entropy problem (since this will be used for the SAM estimation), the entropy difference equation must be changed to account for minimizing the entropy difference of

$$L = \sum_{i=1}^{n} p_i \ln(p_i / q_i) + \sum_{t=1}^{T} \lambda_t [y_t - \sum_{i=1}^{n} p_i \quad f_t(x_i)] + \mu(1 - \sum_{i=1}^{n} p_i) + \sum_{i=1}^{n} \sum_{w=1}^{m} W_{i,w} \ln(\frac{W}{1}$$

the errors as well:[14]

This equation is still subject to moment consistency constraints and additional adding-up constraints, as well as the constraint that the weights sum up to one. As the equation is written here, the terms with the probabilities, **p**, and with the error term, **W**, are equally weighted, reflecting an equal preference for "precision" of the estimates of the parameters (the p_i) and "prediction" of the equation (the W_i). Golan, Judge, and Miller (1996) report Monte Carlo experiments where they explore the implications of changing these weights and conclude that equal weighting of precision and prediction is reasonable.

D. Cross Entropy Application to Balancing SAM

[14] Notice that here, the prior for the cross entropy distance of the W_i is the uniform distribution (1/n), reflecting that there is no prior information.

In applying the entropy technique to a SAM balancing problem, it is helpful to see the relationship between the p's of the information theory problem and the A_{ij}'s in the SAM. Both parameters are constrained by adding up conditions: the p's must sum to one, as the A_{ij}'s must sum to one by column. The moment consistency condition of the SAM is that the column coefficients multiplied by their respective column totals should equal their respective row totals. Usually, some prior coefficient matrix of the SAM is known, so the problem is to minimize the entropy difference between that prior matrix, $\bar{A}$, and the new estimated one, A, subject to the summing up constraints and any other knowledge. If there is measurement error in the data, as expected in the Mexico SAM, the stochastic

$$\min \sum_i \sum_j A_{ij} \left(\ln A_{ij} - \ln \bar{A}_{ij}\right) + \sum_i \sum_w W_{i,w} \ln\left(\frac{W_{i,w}}{1/n}\right)$$

specification can be incorporated.[15] Thus the problem is as follows:

$$\sum_j A_{ij} \; y*_j = y*_i$$

subject to

$$\sum_j A_{ji} = 1$$

and

and the $W_{i,w}$ sum to one across the sectors. The y* are the new row and column sums. Additional information can be added to the constraint equations, in the form of other

[15] In this case, the column totals are assumed to be measured with error.

linear adding-up constraints, as well as through inequalities, if exact relationships are unknown, but upper or lower bounds can be specified.

In the case of the Mexican SAM, the “prior” matrix will not actually be a SAM from an earlier year but rather the proto-SAM described in Section 3. The constraints will include making sure that the micro SAM is consistent with most of the aggregate totals from the macro SAM.[16] Also, the manipulation of the disaggregated agricultural and livestock data will have to be reconciled with the national accounts data. For example, each agricultural sector's activities column is composed of the breakdown between value added and intermediate demand, based on the input-output coefficients from 1985. While this makes the total sum of agricultural gross output equal to the national accounts' sum, the total sum of agriculture value added and the total sum of agricultural intermediate demand do not equal their value added totals from the national accounts. This is because the ratios of value added to intermediate demand must have changed since 1985. However, the national accounts data is assumed to be the more accurate (and current) data, so one restriction will be that the sum of the individual agriculture value added equals the national accounts value added data, and similarly for the intermediate demand. Table 6 contains the specific constraints of the entropy model.

E. Final Steps

The entropy problem for balancing the micro SAM is written into GAMS (General Algebraic Modeling) code[17] according to the equations listed above. The program consists of nearly 8000 single equations, and took about 15 hours to solve. To speed up the solving time, the problem was implemented in an algorithm (developed by Mike Ferris and others at the University of Wisconsin) which converted the equations into their first order conditions. This effectively solves the problem in its dual, which as previously

---

[16] Recall that the payments to factors are manually adjusted in the construction of the proto-SAM in order to diminish the capital intensity of the economy. Thus it would be incorrect to set the factor account aggregates to their respective aggregates in the macro SAM. On the other hand, the *total* value added from the macro SAM should be preserved.

[17] See Brooke, Kendrick and Meeraus (1988), for an introduction to GAMS.

mentioned, is much more efficient from a computational viewpoint. Indeed, with this specification, the micro SAM was balanced in less than 10 *minutes*.

After the micro SAM is balanced, the macro SAM must be recalculated by aggregating up the micro SAM. Table 7 presents the new macro SAM. Since in some cases flows have been adjusted from the original National Accounts data, the new macro SAM will differ from the original one. For example, the proportions of labor and capital use have been altered dramatically, such that the new total for labor value-added more than doubles, while the new total for capital value-added decreases by almost half. These

changes filtered through the rest of the SAM, in terms of factor payments to the households and enterprises. The new macro SAM also accounts for home consumption, which was not in the original macro SAM. Within the micro SAM, most of the adjustment took place in the intermediate demand, which is expected, given the old IO coefficients used. In addition, some changes took place in sectoral investment, since these entries were less constrained than other commodity entries, and thus had to bear the burden of the row-column discrepancies.

Table 6. Specific Constraints of Entropy Model[a]

1. Total Intermediate Demand = Intermediate Demand in Macro SAM
2. Total Exports = Exports in Macro SAM
3. Total Imports = Imports in Macro SAM
4. Total Payment to Land = Payment to Land in Macro SAM
5. Total Value Added = Sum of Payments to Labor, Capital and Land in Macro SAM
6. Total Private Consumption = Private Consumption in Macro SAM, minus Home Consumption from proto-SAM
7. Total Government Consumption = Government Consumption in Macro SAM
8. Total Investment Demand = Investment Demand in Macro SAM
9. Total Activity Tax Revenue = Activity Tax Revenue in Macro SAM
10. Total Commodity Tax Revenue = Commodity Tax Revenue in Macro SAM
11. Total Social Security Tax Revenue = Social Security Tax Revenue in Macro SAM
12. Total Capital Transfers Abroad = Capital Transfers Abroad in Macro SAM
13. Total Income Tax = Income Tax in Macro SAM
14. Total Household Savings = Household Savings in Macro SAM
15. Total Enterprise Tax = Enterprise Tax in Macro SAM
16. Total Enterprise Savings = Enterprise Savings in Macro SAM
17. Total Government Transfers to Household = Government Transfers to Household in Macro SAM (within 5 percent bound)
18. Total Government Payments Abroad = Government Payments Abroad in Macro SAM (Foreign Interest Payments)
19. Total Foreign Transfers to Households = Foreign Transfers to Households in Macro SAM
20. Total Foreign Capital Transfers = Foreign Capital Transfers in Macro SAM

[a]This table shows which micro SAM accounts are constrained in the entropy problem (in addition to the moment consistency constraints and the summing-up condition). For example, line 1 means that the sum of intermediate demands from the proto-SAM must equal the entry in the Macro SAM.

V. Conclusions

This paper has presented the construction of a micro SAM for Mexico, which includes regionalized production, factors and households. Building such a disaggregated SAM requires a large amount of data from a myriad of sources and years, which need to be reconciled with each other. The entropy method is used to provide a consistent and balanced SAM which can then provide the underlying data for further analysis of the Mexican economy.

The changes between the unbalanced proto-SAM and the final micro SAM highlight areas in which the data needs to be improved. Clearly, the SAM construction would benefit from a more up-to-date input-output table. In addition, better data on value-added would eliminate some of the "guess-estimation" in forming the proto-SAM. As mentioned earlier, however, this type of measurement error is endemic in developing countries. Along this vein, more accurate data on household income sources, in particular, from enterprise income, would add to precision of the proto-SAM. Nonetheless, this data-poor environment underscores the need for the entropy method to successfully reconcile the imbalances and inconsistencies of the SAM, given a variety of prior information.

Receipts	Activities	Commodities	Labor	Capital	Land	Households	Enterprise	Government	Savings-Investment	World	Total
Activities		4222				7					4228
Commodities	1856					1689		189	481	790	5004
Labor	1483										1483
Capital	746										746
Land	62										62
Households			1437		62		360	127		6	1992
Enterprise				639				44			683
Government	82	55	46			69	141				393
Savings-Investment						227	182	-12		84	481
World		728		107				45			880

Table .7-- Cont'd

Receipts	Activities	Commodities	Labor	Capital	Land	Households	Enterprise	Governmen t	Savings-Investment	World	Total
Activities		4222				7					4228
Commodities	1856					1689		189	481	790	5004
Labor	1483										1483
Capital	746										746
Land	62										62
Households			1437		62		360	127		6	1992

Enterprise				639				44			683
Government	82	55	46			69	141				393
Savings-Investment						227	182	-12		84	481
World		728		107				45			880
Total	4228	5004	1483	746	62	1992	683	393	481	880	

References

Banco de Mexico. 1997. *The Mexican Economy* (Mexico City).

Brooke, A., De. Kendrick, and A. Meeraus. 1988. *GAMS: A User's Guide,* The Scientific Press, San Francisco.

Golan, Amos, George Judge, and Douglas Miller. 1996. *Maximum Entropy Econometrics, Robust Estimation with Limited Data*. John Wiley & Sons.

Instituto Nacional de Investigaciones Forestales Agricolas y Pecuarias (INIFAP). 1997. *Jornales Por Ha. Por Entidad Federativa.* Mexico (September).

Instituto Nacional de Encuestas, Geografía, e Informática (INEGI). 1994. *Encuesta Nacional de Ingresos y Gastos de los Hogares.*

Instituto Nacional de Encuestas, Geografía, e Informática (INEGI). 1998. *Sistema de Cuentas Nacionales de México.*

Jaynes, E. 1957. Information Theory and Statistical Mechanics. *Physics Review.* 106.

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

Levy, Santiago and Sweder van Wijnbergen. 1992. *Mexican Agriculture in the Free Trade Agreement: Transition Problems in Economic Reform*. OECD Development Centre, Technical Papers. No. 63.

Löfgren, Hans, and Sherman Robinson. 1998. Spatial Networks in Multi-Region Computable General Equilibrium Models. Paper prepared for presentation at the Twelfth International Conference on Input-Output Techniques organized by the International Input-Output Association, New York, May 18-22, 1998.

McDougall, Robert A. 1999. Entropy Theory and RAS are Friends. Paper presented at the Second Annual Conference on Global Economic Analysis, Copenhagen, June.
Norton, Roger and Leopoldo Solis. 1983. *The Book of Chac: Programming Studies for Mexican Agriculture*. Johns Hopkins University Press.

Pyatt, Graham and Jeffery Round, eds. 1985. *Social Accounting Matrices. A Basis for Planning*. Washington, DC: The World Bank.

Robinson, Sherman, Mary Burfisher, and Karen Theirfelder. 1995. *The Impact of the Mexican Crisis on Trade, Agriculture, and Migration*. International Food Policy Research Institute, Trade and Macroeconomics Division, Discussion Paper No. 8.

Robinson, Sherman, Andrea Cattaneo and Moataz El-Said. 1998. *Estimating a Social Accounting Matrix Using Entropy Methods*. International Food Policy Research Institute, Trade and Macroeconomics Division, Discussion Paper No. 33.

Secretaría de Agricultura, Ganadería y Desarrollo Rural (SAGAR). 1996. Anuario Estadístico de la Producción Agrícola de los Estados Unidos Mexicanos, Vols. I and II.

Shannon, C. 1948. A Mathematical Theory of Communciation. *The Bell System Technical Journal*, 27(3).

List of Discussion Papers

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 Lofgren 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)

No. 51 - "Agriculture-Based Development: A SAM Perspective on Central Viet Nam" by Romeo M. Bautista (January 2000)

No. 52 - "Structural Adjustment, Agriculture, and Deforestation in the Sumatera Regional Economy" by Nu Nu San, Hans Löfgren and Sherman Robinson (March 2000)

No. 53 - "Empirical Models, Rules, and Optimization: Turning Positive Economics on its Head" by Andrea Cattaneo and Sherman Robinson (April 2000)

No. 54- "Small Countries and the Case for Regionalism vs. Multilateralism" by Mary E. Burfisher, Sherman Robinson and Karen Thierfelder (May 2000)

No. 55 - "Genetic Engineering and Trade: Panacea or Dilemma for Developing Countries" by Chantal Pohl Nielsen, Sherman Robinson and Karen Thierfelder (May 2000)

No. 56 - "An International, Multi-region General Equilibrium Model of Agricultural Trade Liberalization in the South Mediterranean NIC's, Turkey, and the European Union" by Ali Bayar, Xinshen Diao and A. Erinc Yeldan (May 2000)

No. 57*- "Macroeconomic and Agricultural Reforms in Zimbabwe: Policy Complementarities Toward Equitable Growth" by Romeo M. Bautista and Marcelle Thomas (June 2000)

No. 58 - "Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods " by Sherman Robinson, Andrea Cattaneo and Moataz El-Said (August 2000)

No. 59 - "Food Security and Trade Negotiations in the World Trade Organization : A Cluster Analysis of Country Groups " by Eugenio Diaz-Bonilla, Marcelle Thomas, Andrea Cattaneo and Sherman Robinson (November 2000)

No. 60*- "Why the Poor Care About Partial Versus General Equilibrium Effects Part 1: Methodology and Country Case" by Peter Wobst (November 2000)

No. 61 - "Growth, Distribution and Poverty in Madagascar : Learning from a Microsimulation Model in a General Equilibrium Framework " by Denis Cogneau and Anne-Sophie Robilliard (November 2000)

No. 62- "Farmland Holdings, Crop Planting Structure and Input Usage: An Analysis of China's Agricultural Census" by Xinshen Diao, Yi Zhang and Agapi Somwaru (November 2000)

No. 63- "Rural Labor Migration, Characteristics, and Employment Patterns: A Study Based on China's Agricultural Census" by Francis Tuan, Agapi Somwaru and Xinshen Diao (November 2000)

No. 64 - "GAMS Code for Estimating a Social Accounting Matrix (SAM) Using Cross Entropy (CE) Methods" by Sherman Robinson and Moataz El-Said (December 2000)

No. 65- "A Computable General Equilibrium Analysis of Mexico's Agricultural Policy Reforms" by Rebecca Lee Harris (January 2001)

No. 66- "Distribution and Growth in Latin America in an Era of Structural Reform" by Samuel A. Morley (January 2001)

No. 67 - "What has Happened to Growth in Latin America" by Samuel A. Morley (January 2001)

No. 68 - "China's WTO Accession: Conflicts with Domestic Agricultural Policies and Institutions" by Hunter Colby, Xinshen Diao and Francis Tuan (January 2001)

No. 69 - "A 1998 Social Accounting Matrix for Malawi" by Osten Chulu and Peter Wobst (February 2001)

No. 70 - "A CGE Model for Malawi: Technical Documentation" by Hans Löfgren (February 2001)

No. 71 - "External Shocks and Domestic Poverty Alleviation: Simulations with a CGE Model of Malawi" by Hans Löfgren with Osten Chulu, Osky Sichinga, Franklin Simtowe, Hardwick Tchale, Ralph Tseka and Peter Wobst (February 2001)

No. 72- "Less Poverty in Egypt? Explorations of Alternative Pasts with Lessons for the Future" by Hans Löfgren (February 2001)

No. 73- "Macro Policies and the Food Sector in Bangladesh: A General Equilibrium Analysis" by Marzia Fontana, Peter Wobst and Paul Dorosh (February 2001)

No. 74- "A 1993-94 Social Accounting Matrix with Gender Features for Bangladesh" by Marzia Fontana and Peter Wobst (April 2001)

No. 75- " A Standard Computable General Equilibrium (CGE) Model" by Hans Löfgren, Rebecca Lee Harris and Sherman Robinson (April 2001)

No. 76- " A Regional General Equilibrium Analysis of the Welfare Impact of Cash Transfers: An Analysis of Progresa in Mexico" by David P. Coady and Rebecca Lee Harris (June 2001)

No. 77- "Genetically Modified Foods, Trade, and Developing Countries" by Chantal Pohl Nielsen, Karen Thierfelder and Sherman Robinson (August 2001)

No. 78- "The Impact of Alternative Development Strategies on Growth and Distribution: Simulations with a Dynamic Model for Egypt" by Moataz El-Said, Hans Löfgren and Sherman Robinson (September 2001)

No. 79- "Impact of MFA Phase-Out on the World Economy an Intertemporal, Global General Equilibrium Analysis" by Xinshen Diao and Agapi Somwaru (October 2001)

No. 80*- "Free Trade Agreements and the SADC Economies" by Jeffrey D. Lewis, Sherman Robinson and Karen Thierfelder (November 2001)

No. 81- "WTO, Agriculture, and Developing Countries: A Survey of Issues" by Eugenio Díaz-Bonilla, Sherman Robinson, Marcelle Thomas and Yukitsugu Yanoma (January 2002)

No. 82- "On boxes, contents, and users: Food security and the WTO negotiations" by Eugenio Díaz-Bonilla, Marcelle Thomas and Sherman Robinson (November 2001: Revised July 2002)

No. 83- "Economy-wide effects of El Niño/Southern Oscillation ENSO in Mexico and the role of improved forecasting and technological change" by Rebecca Lee Harris and Sherman Robinson (November 2001)

No. 84- "Land Reform in Zimbabwe: Farm-level Effects and Cost-Benefit Analysis" by Anne-Sophie Robilliard, Crispen Sukume, Yuki Yanoma and Hans Löfgren (December 2001: Revised May 2002)

No. 85- "Developing Country Interest in Agricultural Reforms Under the World Trade Organization" by Xinshen Diao, Terry Roe and Agapi Somwaru (January 2002)

No. 86- "Social Accounting Matrices for Vietnam 1996 and 1997" by Chantal Pohl Nielsen (January 2002)

No. 87- "How China's WTO Accession Affects Rural Economy in the Less- Developed Regions: A Multi-Region, General Equilibrium Analysis" by Xinshen Diao, Shenggen Fan and Xiaobo Zhang (January 2002)

No. 88- "HIV/AIDS and Macroeconomic Prospects for Mozambique: An Initial Assessment" by Channing Arndt (January 2002)

No. 89- "International Spillovers, Productivity Growth and Openness in Thailand: An Intertemporal General Equilbrium Analysis" by Xinshen Diao, Jørn Rattsø and Hildegunn Ekroll Stokke (February 2002)

No. 90- "Scenarios for Trade Integration in the Americas" by Xinshen Diao, Eugenio Díaz-Bonilla and Sherman Robinson (February 2002)

No. 91- "Assessing Impacts of Declines in the World Price of Tobacco on China,Malawi, Turkey, and Zimbabwe" by Xinshen Diao, Sherman Robinson, Marcelle Thomas and Peter Wobst (March 2002)

No. 92*- "The Impact of Domestic and Global Trade Liberalization on Five Southern African Countries" by Peter Wobst (March 2002)

No. 93 - "Trade and tradability: Exports, imports, and factor markets in the Salter-Swan model " by Karen Thierfelder and Sherman Robinson (May 2002)

No. 94- "That was then but this is now: Multifunctionality in industry and agriculture" by Eugenio Díaz-Bonilla (May 2002)

No. 95- "A 1998 social accounting matrix (SAM) for Thailand" by Jennifer Chung-I Li (July 2002)

No. 96- "Trade and the skilled-unskilled wage gap in a model with differentiated goods" by Karen Thierfelder and Sherman Robinson (July 2002)

No. 97- "Estimation of a regionalized Mexican Social Accounting Matrix: Using entropy techniques to reconcile disparate data sources" by Rebecca Lee Harris (September 2002)

TMD Discussion papers marked with an '*' are MERRISA-related. Copies can be obtained by calling Maria Cohan at 202-862-5627 or e-mail: m.cohan@cgiar.org