Reconciliation of the GTAP and Household Survey Data

Maros Ivanic

GTAP Research Memorandum No. 5

February 2004

Abstract

This paper presents a method that was employed in order to make the available fourteen household income survey data sets compatible with the data in the GTAP version 5 database. The first step of the method was the imputation of the unobservable returns to GTAP factors from the reported data. The second step was the reconciliation of the two data sets so that their joint totals would be identical.

The paper does not claim to be the final word on either data imputation or reconciliation; instead it works through the various issues encountered in the process, proposes solutions to them and leaves it to the reader to judge the validity of this methodology.

1 Introduction

The purpose of this paper is to explore and resolve the challenges of merging two very different data sets into a single improved set: the well established, thoroughly reconciled and homogenous GTAP database on one hand, and the rather raw and heterogenous survey data on the other. Both of these data sets possess critical information for assessing the links between trade and poverty, however the differences that exist between these sets make it impossible to use them together,without significantly altering either one or both of them. Any such alteration, of course, means throwing out some information and therefore it is important that this information be either incorrect or at least less reliable than the rest in order for the product to be an improvement over the initial data. Such a proper identification, is of course very difficult and has to rely on a thorough understanding of the links between the data in question and a set of reasonable assumptions that permit ranking the quality of the relevant pieces of information.

Section 2 describes the household survey data in detail and lays out the procedure for extracting the primary factor returns data from the survey data that would make them directly comparable to the GTAP data. This extraction is an exercise of imputation of returns to basic GTAP factors from the profits and wages reported by the households in the survey data. It also serves as a thorough description of the available survey data. The description of the GTAP data set is largely omitted, because it is described in detail elsewhere [2].

The links between the household income data and the GTAP data are implied by the information contained in the respective data sources. The household income data contain, in principle, the level of returns to primary factors in the interviewed sample of households. Thus the survey data is closely related to the information on value-added (payments to primary factors) in the GTAP database (arrays VFM, EVOA and EVFA). Under ideal conditions, it would be natural to expect that the per capita level of factor income should be the same in the GTAP data and in the survey data. Moreover the composition of the returns to all factors should be the same in both data sets. Additionally, if the survey data is disaggregated into the same sectors as exist in the GTAP

database (arrays VFM and EVFA), then the factor usage in the sectors should again be the same in the survey data and in the GTAP tables.

These links are discussed more closely in Section 3, which describes the reconciliation method between the data by evaluating the quality of linked pieces of information. By stating the assumptions on the quality of each data set within each link, it is able to adjust the data by removing the low quality data and thus reconciling the remaining information contained in the sets. This process is done to the point where both of the sets are fully compatible. Section 4 offers some conclusions.

2 Household Survey Data

2.1 Data Description

Fourteen household surveys (Table 1) were used to extract the information on factor composition of household factor earnings. These surveys were chosen from the set of full set of household surveys available to the World Bank in July 2002 based on various criteria. First, the data set had to be available and sufficiently informative. The minimum requirement on data content was that the data contain information on wage and business income, transfers and a sufficient number of characteristics to distinguish the nature of household member's employment and business. The second criterion for inclusion of the household survey was that the country be included in the GTAP v. 5 global database [1] so that results from further analyses could be later applied in the GTAP framework. Finally a country had to have a measurable segment of the population earning less than $1 per day to justify its inclusion in this research.

[1] It should be noted that the set of countries in GTAP is constantly expanding. Version 6 has 20 more regions than Version 5.

Country	**Sample Size**	**Year**	**Name of Survey**
Bangladesh	7,417	1996	Household Expenditure Survey [8]
Brazil	88,972	1998	Pesquisa Nacinlal por Amostra de Domicilios [13]
Chile	47,805	1998	Encuesta de Caracterizatcion Socioeconomica Nacional [7]
Colombia	30,527	1998	Encuesta Nacional de Hogares Fuerza de Trabajo [6]
Indonesia	59,111	1993	National Socio-Economic Survey [14]
Malawi	9,243	1998	Integrated Household Survey [11]
Mexico	10,072	2000	Encuesta Nacional de Ingresos y Gastos de los Hogares [9]
Peru	6,732	1999	Encuesta nacional de Hogares [4]
Philippines	37,393	1999	Annual Poverty Indicators Survey [3]
Thailand	25,106	1996	Socio Economic Survey [15]
Uganda	10,680	1999	Uganda National Household Survey [16]
Venezuela	16,022	1998	Encuesta de Hogares por Muestreo [5]
Vietnam	5,999	1998	Household Living Standards Survey [10]
Zambia	15,268	1999	Living Conditions Monitoring Survey [12]

Table 1: Household surveys used in the study.

2.2 Observed Income in Household Surveys

For the purposes of household surveys, the heads of households were supposed to report all actual income that was earned by any of the household members, including all sorts of transfers, employee benefits and in-kind payments and excluding all income that never materialized. This means that non-monetary or irregular benefits, such as free rent, tax refunds and lottery winnings were included in household income, while wages in arrears were not. Because income was reported as it was made available to the household, it could, in some cases, represent the disposable income net of income taxes, however in most cases it represented household's gross income.

The assumption that the reported income includes all household income should be viewed with suspicion: even though the households were asked to be honest in reporting all their incomes (with granted confidentiality from the collecting agencies) it is likely that some households, despite the assurances, did not report all of their income, resulting in underreporting the total household income in these cases.

The set of primary factors considered in this study consisted of the following: skilled labor (l_s), unskilled labor (l_u), capital used in agriculture (c_a), capital used in non-agriculture (c_n), land (c_l) and transfers (t). These are directly relatable to the following GTAP factors (endowments): skilled labor, unskilled labor, land, capital in GTAP agricultural sectors [2], and capital in non-agricultural sectors.

Some of the income sources reported in household surveys can be directly linked to these factors. For example, wage labor is a subset of the returns to labor. On the other hand, other income sources represent apparent profits to households that are usually without a clear connection to the underlying factors that earned these profits. For that reason, reported income had to be processed to determine factor content of each reported source of a household's income. The process of allocating factor content of reported income was done by mapping reported income categories into primary factors under various assumptions, which are described below; for the formal description of the assumptions used in deriving factor income from the primary variables please see Table 6 and Figure 1.

All sources of income observed in household surveys were combined into a condensed set of five variables, including two variables capturing reported wages, further distinguished into skilled and unskilled wages (w_s,w_u) by the occupation variable; two variables for business income distinguished into agricultural (b_a) and nonagricultural (b_n) by nature of the business; one variable for transfers received by household from private and public (government) sources (t) as well as variables collecting personal characteristics on household members such as age, education, industry of employment and skill level. These personal variables

[2] Agricultural sectors are defined as including: PDR Paddy rice, WHT Wheat, GRO Cereal grains nec, V_F Vegetables, fruit, nuts, OSD Oil seeds, C_B Sugar cane, sugar beet, PFB Plant-based fibers, OCR Crops nec, CTL Bovine cattle, sheep and goats, horses, OAP Animal products nec, RMK Raw milk, WOL Wool, silk-worm cocoons, FOR Forestry, FSH Fishing

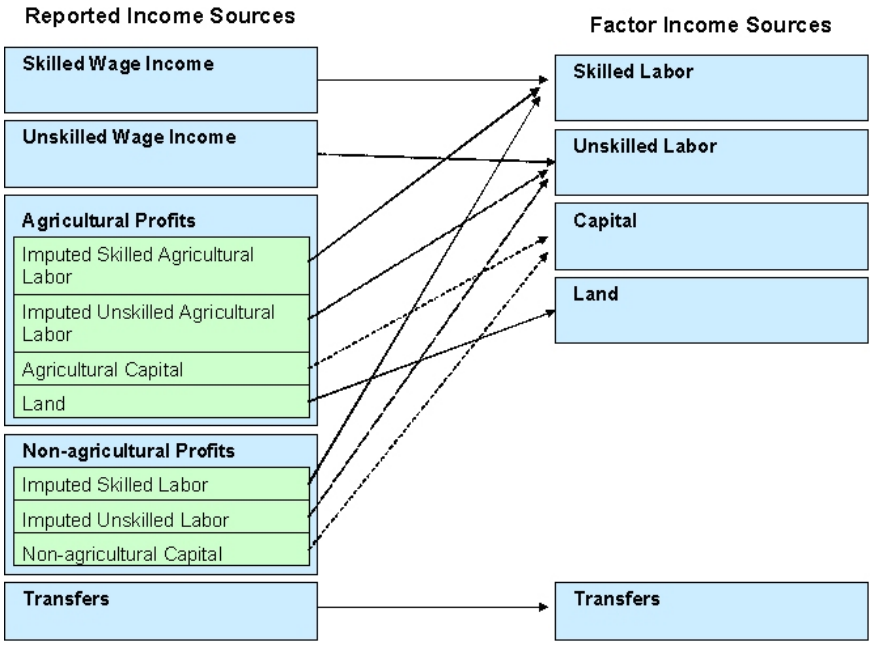

Figure 1: Structure of Factor Income Imputation.

were taken directly from the surveys, with the exception of skill level of a worker that was determined by the person's occupation whereas all professionals and managers were considered skilled and the others unskilled, following the GTAP definition of skilled labor. For the list of the primary variables with descriptions see Table 2.

2.3 Derivation of primary factor income from reported income

The process of linking primary factors to the income observed in the primary variables was based on a set of assumptions that mapped each reported income to primary factors and that were globally applied to each survey:

Assumption 1 (Wage labor) *Wage income includes all income (cash and in-kind), which was received by a household member in return for his/her labor and the receiver was not an employer or self-employed.*

Assumption 2 (Skilled labor) *The skill level of labor was determined by the GTAP definition: all professionals and managers were classified as skilled, all others as unskilled. (This definition applies to both reported wages and imputed returns to labor.)*

Income Variable	**Description**
w_s	Reported skilled wage income
w_u	Reported unskilled wage income
$\bar{w}_s$	Imputed Skilled Wage
$\bar{w}_u$	Imputed Unskilled Wage
$\bar{w}_a$	Imputed Wage in Agricultural Business Income
$\bar{w}_n$	Imputed Wage in Non-Agricultural Business Income
b_a	Reported income of households agricultural business
b_n	Reported income of households nonagricultural business
p_a	Reported rent income from households agricultural property (land, agricultural equipment)
p_n	Reported rent income from households nonagricultural property (house rents, dividends)
t	Private and pubic (government) transfer income
Characteristics Variable	
IND	Industry
OCU	Skill level (0 for unskilled, 1 for skilled)
ALE	Age level
ELE	Education level
Factor Income	
l_s	Skilled Labor
l_u	Unskilled Labor
c_a	Return to Agricultural Capital
c_n	Return to Non-Agricultural Capital
c_l	Return to Land
t	Transfers

Table 2: Variables defined for the purpose of factor income extraction from household surveys

The imputed and wage income, before they could be added to the respective returns to labor, were classified based on a person's skill level as either skilled (w_s, $\bar{w}_s$) or unskilled (w_u, $\bar{w}_u$).

Assumption 3 (Transfers) *Reported government and private transfers represent transfer income.*

The assumptions that governed sources of reported agricultural (b_a) and non-agricultural (b_n) business income, a potential blend of returns to all capital, labor and land, were more involved:

Assumption 4 (Property rents) *The returns to capital whose source was clearly identifiable (such as dividends, property rents i.e.) represent capital returns. These were further split according to its reported nature* [3] *to either agricultural capital return or nonagricultural capital return.*

Profits would not be complete without property rents directly reported by the households. Therefore, the reported agricultural (p_a) and non-agricultural (p_n) property income was added to their respective types of capital, or $p_a \rightarrow c_a$ and $p_n \rightarrow c_n$.

Assumption 5 (Agriculture vs. non-agriculture) *The definition of agriculture and non-agriculture came from the definitions in each household survey. Agriculture included forestry and fishing.*

Assumption 6 (Imputation of capital) *The value of returns to capital for a household is the residual of the reported profits less the estimated return to labor.*

Assumption 7 (Classification of sector in the absence of data) *The agricultural or non-agricultural nature of wage and imputed labor returns for household members without clear job description was determined by the occupation of the household head.*

Because of significantly less information on returns to capital as compared to reported returns to wage labor, the value of returns to capital was calculated as the residual to the reported profits less the estimated return to labor, which was calculated as imputed labor income ($\bar{w}$) for all household members involved in the business. This imputed income for a household member was determined as the average wage of all workers in the economy that earned wage income only and possessed an identical set of personal characteristics of age, education level, skill level, and industry of employment.

Imputed labor income was then subtracted from the reported income of the household businesses; to improve accuracy, this was done keeping agricultural

[3] Agricultural property rents include rental payments for land, farming equipment and the similar objects used in agriculture. All other property rents were classified as non-agricultural property rents.

Country	**Share of Land in Agricultural Capital**	**Country**	**Share of Land in Agricultural Capital**
Bangladesh	0.464	Peru	0.402
Brazil	0.201	Philippines	0.545
Chile	0.388	Thailand	0.46
Colombia	0.488	Uganda	0.358
Indonesia	0.607	Venezuela	0.439
Malawi	0.334	Vietnam	0.676
Mexico	0.446	Zambia	0.271

Table 3: Applied land cost shares in agricultural production

and non-agricultural business income separate so that only imputed agricultural wage was subtracted from the agricultural profits, and non-agricultural imputed wage was only subtracted from the non-agricultural profits.

Any residual left was then ascribed to return to nonagricultural capital (c_n) if the business was in non-agricultural sector, or a sum of combined returns to agricultural capital (c_a) and land (c_l) if the business was engaged in agricultural activities. If the total of imputed wage was greater than the total of reported business income, then all operating surplus was classified as return to labor and capital content of this business was set to zero. Symbolically, the process of breaking up business income can be written as: $b_a \rightarrow \bar{w}_a + c_a + c_l$, $b_n \rightarrow \bar{w}_n + c_n$

Assumption 8 (Returns to Land) *The value of returns to land was determined as a fixed ratio of capital returns in agriculture, obtained from the GTAP database.*

The portion of returns to land in residual sum of returns to capital and land for agricultural businesses was not estimated from the datasets because of insufficient information to allow imputation of returns to land in most of surveys. Instead it was based on outside information (GTAP database) that specified the portion of payments to land within agricultural sector of each country (α) and this ratio was then uniformly applied to the sum of agricultural capital and land derived in the step before. The applied shares are reported in Table 3.

	Reported Agricultural Business Income=100	**Reported Nonagricultural Business Income=50**
(1) Reported Business Income	70+30=100	50+0=50
(2) Imputed Wage	70	50
(3) Residual (1)-(2)	100-70=30	50-50=0
(4) Reported Property Income	20	50
(5) Land+Capital (2)+(4)	20+30=50	0
(6) To Land (5)α	50α	0
(7) To Capital (5)(1-α)	50(1-α)	0

Table 4: Example of Breaking up Business Income

2.4 Example

Working through an example can be a good way to illustrate the whole process of deriving factor incomes from reported business income. Let's assume that a household reported agricultural business income of 100, agricultural property income of 20, and nonagricultural business income of 50. Let's further assume that, based on the characteristics of household members employed in these businesses, they were ascribed 70 of imputed labor income in agricultural activities and 50 of imputed labor income in nonagricultural activities. The combined return to land and capital for the agricultural enterprise would then be the profit of 100 less the imputed wage of 70 plus the agricultural rents of 20 for the total of 50. Given the share level of land in total returns to land and capital in the country α, return to land can be further separated as 50 and return to capital as $50(1-\alpha)$. In the nonagricultural business the return to capital would be zero, because the imputed wages equal the reported income. Table 4 presents this treatment in a more schematic way.

2.5 Results and Discussion

Table 5 contains factor income shares for the fourteen countries as obtained from the household survey data.

Table 5 prompts a few thoughts on the quality of the derived data. First, we notice a rather uniform pattern of factor shares across countries: (1) the skilled wage share is almost always lower than the share of unskilled wage, (2) much less of skilled labor is imputed than of unskilled labor, and (3) the share of returns to land is generally low. This may suggest that the data are indeed not random and therefore the reported shares may represent the true shares. Judging the quality of the respective data for each factor income category is more difficult.

	Skilled Wage	**Unskl Wage**	**Impt'd Skilled Labor**	**Impt'd Unskl Labor**	**Capital**	**Land**	**Total**
Bangladesh	0.10	0.24	0.00	0.46	0.17	0.03	1.00
Brazil	0.29	0.30	0.08	0.09	0.22	0.01	1.00
Chile	0.17	0.33	0.03	0.08	0.34	0.04	1.00
Colombia	0.18	0.37	0.03	0.27	0.13	0.01	1.00
Indonesia	0.18	0.24	0.02	0.42	0.11	0.02	1.00
Malawi	0.29	0.20	0.02	0.20	0.26	0.03	1.00
Mexico	0.20	0.49	0.03	0.16	0.11	0.01	1.00
Peru	0.28	0.28	0.08	0.25	0.09	0.03	1.00
Philippines	0.24	0.34	0.02	0.10	0.22	0.08	1.00
Thailand	0.20	0.34	0.03	0.24	0.16	0.02	1.00
Uganda	0.11	0.16	0.01	0.31	0.33	0.09	1.00
Venezuela	0.13	0.25	0.03	0.28	0.29	0.02	1.00
Vietnam	0.02	0.31	0.02	0.40	0.20	0.05	1.00
Zambia	0.16	0.33	0.01	0.23	0.27	0.00	1.00

Table 5: Household survey results: national factor income shares

Factor return	**Definition**
Skilled labor	$l_s = w_s + \bar{w}_s$
Unskilled labor	$l_u = w_u + \bar{w}_u$
Ag. capital	$c_a = \max(0, (1-\alpha)(b_a - \bar{w}_a)) + p_a$
Non-ag. capital	$c_n = \max(0, (b_n - \bar{w}_n)) + p_n$
Land	$c_l = \max 0, \alpha(b_a - \bar{w}_a)$
Transfers	t
	where α is the GTAP determined land share of comprehensive agricultural capital

Table 6: Definition of factor income in household surveys.

The household data look least reasonable in the column with the share of capital returns, which appears to be underreported in many cases: 9% in Peru, 11% in Mexico and Indonesia, 13% Colombia and 16% Thailand. Therefore a natural question arises whether/how we could proceed to improve these data using some other data sources. Section 3 will provide additional discussion on the methods improving the household data with the GTAP data.

	Land	Unskilled Labo	Skilled Labor	Capital	Natural Resources	Total
Bangladesh	0.06	0.39	0.11	0.41	0.02	1.00
Brazil	0.01	0.34	0.16	0.48	0.00	1.00
Chile	0.02	0.28	0.11	0.57	0.01	1.00
Colombia	0.03	0.37	0.16	0.42	0.02	1.00
Indonesia	0.08	0.34	0.07	0.48	0.03	1.00
Malawi	0.04	0.43	0.09	0.43	0.01	1.00
Mexico	0.03	0.23	0.09	0.64	0.02	1.00
Peru	0.03	0.20	0.10	0.66	0.01	1.00
Philippines	0.06	0.32	0.11	0.48	0.02	1.00
Thailand	0.03	0.12	0.04	0.80	0.01	1.00
Uganda	0.06	0.48	0.07	0.38	0.01	1.00
Venezuela	0.02	0.27	0.11	0.54	0.06	1.00
Vietnam	0.06	0.33	0.09	0.49	0.03	1.00
Zambia	0.03	0.40	0.10	0.46	0.01	1.00

Table 7: Original GTAP factor income shares (from EVOA)

2.6 Factor Income Information in the GTAP database

Information on factor income is contained in the GTAP database under headers EVOA [4], EVFA [5] and VFM [6]. Instead of providing more information on the source of the data for each country, the reader is here referred to the documentation of the GTAP database [1]. The overview of the data contained in the GTAP database are in Table 7.

Some differences between the GTAP data and the data derived from household surveys become apparent following the inspection of Tables 7 and 5: capital share in the GTAP data is much higher for every country than it it in the survey data. Because the combined shares of land and natural resources in GTAP are quite similar to the land share in the survey data, it seems that the missing capital in survey data is probably included in the labor share. Because in the survey data labor is broken into wage and imputed labor, we can compare the labor share from GTAP with the wage labor in the survey: we see that these two shares are far more comparable, meaning that the main difference between Tables 7 and 5 appears to be in imputed labor. This is not surprising, since the imputation of labor is one of the most difficult aspects of preparing a national input-output table of the sort used by GTAP.

[4] Endowments - Output at Agents' Prices
[5] Endowments - Firms' Purchases at Agents' Prices
[6] Endowments - Firms' Purchases at Market Prices

3 Reconciliation

3.1 Initial Comparison of the Datasets

The initial inspection of the data sets showed significant differences between the GTAP and survey data on factor income. As with any empirical work, these differences are largely anticipated because of various reasons. First, the surveys were not undertaken in the same time period when the GTAP data were collected. While the GTAP database represents the world in 1997, the surveys were taken in various years between 1993 and 2000 (Table 1). Second, the nature of the data is different in surveys than in the GTAP database. While household surveys contain information that was aggregated from the bottom to the top without placing constraints on the totals, in the GTAP database the information came from aggregate national accounts that were broken down into social accounting matrices in order to sustain the basic theory and accounting principles. Because of the differences in the data collection methodologies, both data sources are likely to suffer from different deficiencies. However, through a sensible reconciliation of the survey data with the GTAP database, a data set superior to both of them can be obtained.

The process of reconciliation was naturally greatly determined by the author's perception of the quality and strengths of the respective data sources. Though every effort was made to base this judgement on the generally accepted properties of either data set, some degree of subjectivity could not be ruled out. Alternative approaches are definitely permissible. The following approach is only one of many other possible. However, it possesses an important merit of simplicity and straightforwardness. It rests upon three basic links identified between the two data sources. The first link is that the per capita income in the GTAP data should be approximately the same as the per capita income implied by the survey data [7]. The second link is that the shares of factor incomes in the GTAP data and the household survey should be the same and, finally, that the sectoral use of endowments (agriculture/non-agriculture) should be the same. Naturally, for the reasons outlined before, the two data sets always differ along these three dimensions.In order to resolve these data conflicts, a set of assumptions had to be adopted in order to govern the primacy of data sources in order to decide on the accepted values.

The two main assumptions governing the reconciliation may be summed up in two statements that are believed to be true in the author's view. The first belief is that the GTAP data base reflects better than the household survey data the total value added for each country in the main sectors of the economy. That means that the total returns to factors employed in the agricultural and non-agricultural sectors in each country were believed to be better represented by the aggregate GTAP numbers than by the survey. There are various reasons for this belief. First, it is very likely that surveys miss the richest households, increasing the chances of underreporting the size of economic activity. This is

[7] One natural source of discrepancies could be caused by the exclusion of taxes on factor returns for some returns in some surveys.

also evidence of underreporting of income by the wealthiest households [17].This then implies that the the totals for the respective sectors and the whole may be wrong.

The second belief is that the survey data reflects better the composition of returns to the easily observable factors, such as skilled and unskilled labor in both agricultural and non-agricultural sectors. This belief is substantiated by the fact that the surveys tend to do a good job in capturing and classifying workers and that there is no significant bias caused to that effect by the problems in sampling and richer households' refusal to participate.

These two assumptions may be easily tested, because they jointly imply that the volume of value-added implied by the survey is strictly lower than that in the GTAP data. If this is true, then the positive difference in the data set values should equal the level of underreported agricultural and non-agricultural capital in each country and this information could be added to the survey data to complete the reconciliation. Table 8 indeed supports this hypothesis by listing the observed values of value-added as contained in the GTAP and household survey data. Note that for each country the amount of value added is strictly lower in the survey data than in the GTAP database.

Because the belief that the survey data underreports the value added appears to be well confirmed by the data, the resulting method of reconciliation was based largely on this fact. This meant that each household survey was brought to the level of the GTAP totals by adjusting the level of imputed agricultural and non-agricultural capital. In all cases of non-agricultural capital and most cases of agricultural capital, this treatment resulted in increasing the level of capital. Only in four cases, some of the agricultural capital was reclassified as non-agricultural capital when the household survey reported more agricultural capital than the GTAP data. This again followed the assumed supremacy of the GTAP totals. The final adjusted data are listed in tables 11 to 24.

The adjustments, at this point, are only aggregate and need to be allocated to the respective households in order to obtain a fully consistent set of household data. Because no other information was available in order to ascribe adjustments to particular households, all adjustments were shared by households relative to their income. This meant, for example, that if the agricultural capital level was increased by 5%, each household's agricultural capital was raised by the same 5%.

Though the aggregate GTAP value-added totals did not change in the process of reconciliation, the composition of the components did, because the information on the shares of most primary factors was taken directly from the survey data. These changes had to be implemented to three endowment tables in the GTAP database: EVOA [8], EVFA [9] and VFM [10]. These changes were implemented as follows. The EVOA table, which is a $REGIONS \times ENDOWMENTS$ table was changed directly to represent the shares suggested by the adjusted survey data. Introducing this information to the table did not

[8] Endowments - Output at Agents' Prices
[9] Endowments - Firms' Purchases at Agents' Prices
[10] Endowments - Firms' Purchases at Market Prices

	GTAP VA	Survey VA	Difference
Bangladesh	41,469	27,499	13,970
Brazil	729,923	379,334	350,589
Chile	66,431	33,616	32,815
Colombia	87,415	42,539	44,876
Indonesia	206,360	58,941	147,419
Malawi	2,565	1,131	1,433
Mexico	345,231	135,960	209,271
Peru	58,193	15,233	42,961
Philippines	69,868	34,918	34,950
Thailand	140,923	76,367	64,556
Uganda	6,638	3,948	2,691
Venezuela	79,341	27,979	51,362
Vietnam	18,456	16,074	2,383
Zambia	3,759	1,973	1,785

Table 8: Comparison of returns to factor (value-added) in the GTAP database and the household survey data. The values are given in millions of USD. Each country's survey data were converted by the average exchange rate for 1997.

produce any inconsistencies within it or the database as total income was left unchanged. The EVFA and VFM tables, however contain additional information on the use of the endowments in specific production sectors and they are of the size $REGIONS \times ENDOWMENTS \times SECTORS$. Because the introduced information only affected the summations of endowments across two broad (agricultural and non-agricultural) sectors, I used the RAS procedure (explained more closely in the appendix) to adjust each individual cell within the matrix so that the endowment summations and the production summations remained the same. Though this procedure is mathematically not guaranteed to work in all cases, in these fourteen countries it produced a reasonable solution.

4 Conclusion

The method described in this paper permits us to construct a viable compromise of two very different data sets. Though this method is based on various subjective assumptions and thus is limited in the general applications, hopefully it could serve as one of the building steps in the future work on reconciling microeconomic data with national accounts data.

	Agricultural Sector		
	GTAP	Survey	Adjustment
Bangladesh	7,947	9,012	-1,065
Brazil	68,912	60,731	8,181
Chile	6,903	7,607	-704
Colombia	10,846	4,016	6,830
Indonesia	40,536	16,632	23,904
Malawi	747	357	390
Mexico	35,087	15,030	20,057
Peru	6,293	6,741	-447
Philippines	15,315	7,431	7,884
Thailand	14,561	13,483	1,078
Uganda	3,670	1,873	1,797
Venezuela	5,766	2,108	3,658
Vietnam	2,676	2,930	-254
Zambia	951	107	844

Table 9: Adjustments in agricultural sector. In millions of USD valued in 1997.

	Non-Agricultural Sector		
	GTAP	SURVEY	Adjustment
Bangladesh	33,522	18,488	15,035
Brazil	661,012	318,603	342,408
Chile	59,528	26,009	33,519
Colombia	76,569	38,523	38,046
Indonesia	165,824	42,310	123,514
Malawi	1,817	774	1,043
Mexico	310,144	120,930	189,214
Peru	51,900	8,492	43,408
Philippines	54,553	27,487	27,067
Thailand	126,361	62,884	63,478
Uganda	2,969	2,075	893
Venezuela	73,575	25,871	47,704
Vietnam	15,780	13,143	2,637
Zambia	2,807	1,866	941

Table 10: Adjustments in non-agricultural sector. In millions of USD valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	145	-
Skilled Wages	106	2,663
Unskilled Wages	2,530	4,062
Imputed Skilled Labor	6	130
Imputed Unskilled Labor	4,996	7,935
Capital	163	18,732

Table 11: Reconciled data for Bangladesh. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	4,267	-
Skilled Wages	14,774	96,384
Unskilled Wages	22,562	90,273
Imputed Skilled Labor	5,180	26,579
Imputed Unskilled Labor	5,168	30,720
Capital	16,961	417,055

Table 12: Reconciled data for Brazil. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	1,183	-
Skilled Wages	238	5,486
Unskilled Wages	2,578	8,621
Imputed Skilled Labor	59	1,041
Imputed Unskilled Labor	977	1,753
Capital	1,867	42,627

Table 13: Reconciled data for Chile. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	3,731	-
Skilled Wages	166	7,609
Unskilled Wages	742	15,183
Imputed Skilled Labor	12	1,375
Imputed Unskilled Labor	2,453	9,124
Capital	3,742	43,277

Table 14: Reconciled data for Colombia. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	16,928	-
Skilled Wages	136	10,440
Unskilled Wages	2,312	11,990
Imputed Skilled Labor	190	1,227
Imputed Unskilled Labor	11,625	13,211
Capital	9,344	128,955

Table 15: Reconciled data for Indonesia. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	170	-
Skilled Wages	20	304
Unskilled Wages	48	183
Imputed Skilled Labor	1	17
Imputed Unskilled Labor	173	49
Capital	336	1,264

Table 16: Reconciled data for Malawi. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	10,314	-
Skilled Wages	701	26,350
Unskilled Wages	7,228	59,530
Imputed Skilled Labor	43	4,324
Imputed Unskilled Labor	4,119	17,872
Capital	12,682	202,067

Table 17: Reconciled data for Mexico. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	241	-
Skilled Wages	954	3,261
Unskilled Wages	2,151	2,106
Imputed Skilled Labor	332	855
Imputed Unskilled Labor	2,304	1,515
Capital	311	44,162

Table 18: Reconciled data for Peru. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	6,717	-
Skilled Wages	2,682	5,866
Unskilled Wages	125	12,073
Imputed Skilled Labor	313	254
Imputed Unskilled Labor	13	3,647
Capital	5,465	32,713

Table 19: Reconciled data for Philippines. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	2,327	-
Skilled Wages	354	15,009
Unskilled Wages	2,780	23,531
Imputed Skilled Labor	796	1,227
Imputed Unskilled Labor	5,739	12,737
Capital	2,566	73,857

Table 20: Reconciled data for Thailand. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	980	-
Skilled Wages	24	408
Unskilled Wages	87	549
Imputed Skilled Labor	1	45
Imputed Unskilled Labor	821	391
Capital	1,757	1,577

Table 21: Reconciled data for Uganda. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	1,891	-
Skilled Wages	13	5,599
Unskilled Wages	742	10,253
Imputed Skilled Labor	8	904
Imputed Unskilled Labor	845	6,861
Capital	2,268	49,958

Table 22: Reconciled data for Venezuela. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	211	-
Skilled Wages	38	1,610
Unskilled Wages	597	2,708
Imputed Skilled Labor	2	77
Imputed Unskilled Labor	1,727	3,692
Capital	100	7,693

Table 23: Reconciled data for Vietnam. Value-added in millions of US dollars valued in 1997.

	Agricultural Sector	Non-Agricultural Sector
Land	241	-
Skilled Wages	8	307
Unskilled Wages	35	607
Imputed Skilled Labor	0	29
Imputed Unskilled Labor	29	419
Capital	637	1,445

Table 24: Reconciled data for Zambia. Value-added in millions of US dollars valued in 1997.

A The RAS-procedure Used to Update GTAP Endowment-Usage Matrices

Two tables in the GTAP database (VFM and EVFA) required a more involved procedure in order to bring the endowment totals to the level suggested by the adjusted survey data. This was caused by the fact that these matrices are three-dimensional ($REGIONS \times ENDOWMENTS \times SECTORS$), meaning that adjusting a particular endowment total required adjusting a whole string of cells (in sectors) without changing the sectoral total. Because of such a nature of the problem, where a multiple totals are to be enforced at the same time, a modified iterative RAS-procedure was used.

The definition of the problem in the reconciliation setting was following. An original matrix of endowment usage per sectors per country M was adjusted so that its sectoral endowment usage would be identical to that in the household surveys (set of subtotals $T = \{t_1, t_2, t_3, ...\}$), while the total endowment usages per sector in each country remained unchanged (set of subtotals $S = \{s_{l_s}, s_{l_u}, s_{c_a}, s_{c_n}, s_{c_l}\}$) and the original data in M remained as little distorted as possible. Thus we are looking for a matrix M' whose subtotals $i \in S$ are equal to the original matrix M subtotals $s_i(M') = s_i(M)$, while the extraneously given subtotals $j \in T$ are imposed from from the outside data $s_j(M') = G_j$. Tables 25 and 26 illustrate the organization of subtotals.

Country X	Agricultural Sectors			Non-Agricultural Sector		
	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
Skilled Labor	s_{l_s}	s_{l_s}	s_{l_s}	s_{l_s}	s_{l_s}	s_{l_s}
Unskl. Labor	s_{l_u}	s_{l_u}	s_{l_u}	s_{l_u}	s_{l_u}	s_{l_u}
Capital	s_{c_a}	s_{c_a}	s_{c_a}	s_{c_n}	s_{c_n}	s_{c_n}
Nat. Res.	s_{c_a}	s_{c_a}	s_{c_a}	s_{c_n}	s_{c_n}	s_{c_n}
Land	s_{c_l}	s_{c_l}	s_{c_l}	s_{c_l}	s_{c_l}	s_{c_l}

Table 25: The assignment of endowment subtotals within the GTAP endowment usage matrices.

Country X	Agricultural Sectors			Non-Agricultural Sector		
	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
Skilled Labor	t_1	t_2	t_3	t_4	t_5	t_6
Unskl. Labor	t_1	t_2	t_3	t_4	t_5	t_6
Capital	t_1	t_2	t_3	t_4	t_5	t_6
Nat. Res.	t_1	t_2	t_3	t_4	t_5	t_6
Land	t_1	t_2	t_3	t_4	t_5	t_6

Table 26: The assignment of sectoral subtotals within the GTAP endowment usage matrices.

The procedure is standardly carried out through an iterative, two-step process. In the first step, all sub-cells belonging to the subtotals of M that are to be changed ($j \in T$)are multiplied by a scaling factor that brings the given subtotal to the desired level $s_j(M') = G_j$. Of course, after this multiplication, the original condition $[s_i(M') = s_i(M)]$ generally will not hold. This is corrected in the second step where the cells belonging to subtotals $i \in S$ are similarly scaled so that the new matrix M'' is such that $s_i(M) = s_i(M'')$. This adjustment again distorts subtotals $j \in T$, however, in general, the distortion will be smaller than it was before step one. At this point M may be replaced with M'' and the steps repeated until M' converges sufficiently to M''.

References

[1] Dimaranan, Betina V., Robert A. McDougall, Global Trade, Assistance, and Production: The GTAP 5 Data Base, Center for Global Trade Analysis, Purdue University, 2002

[2] Hetel, Thomas W. (editor), Global Trade Analysis: Modeling and Applications, Cambridge University Press, 1997

[3] Annual Poverty Indicators Survey, National Statistics Office, Manila, Philippines, 1999

[4] Encuesta nacional de Hogares, MECOVI, Peru, 1999

[5] Encuesta de Hogares por Muestreo, BID-BIRD-CEPAL, Venezuela, 1998

[6] Encuesta Nacional de Hogares Fuerza de Trabajo, BID-BIRD-CEPAL, Colombia, 1998

[7] Encuesta de Caracterizatcion Socioeconomica Nacional, Ministerio de Planification y Cooperacion, Chile, 1998

[8] Household Expenditure Survey, Bangladesh Bueau of Statistics, Bangladesh, 1996

[9] Encuesta Nacional de Ingresos y Gastos de Los Hogares, Instituto Nacinal de Estadistica Geograpfia e Informatica, Mexico, Mexico, 2000

[10] Household Living Standards Survey, General Statistics Office, Vietnam, 1998

[11] Integrated Household Survey, National Statistical Office, Zomba, Malawi, 1998

[12] Living Conditions Monitoring Survey II (1998) Central Statistical Office, Zambia

[13] Pesquisa Nacional por Amostra de Domicilios (1998), Brazilian Institute of Geography and Statistics (IBGE).

[14] SUSENAS: Indonesia's Socio-Economic Survey (1993) Biro Pusat Statistik, Jakarta, Indonesia.

[15] Thailand Socio-Economic Survey, National Statistics Division, Thailand, 1996

[16] Uganda Bureau of Statistics, National Household Survey 1999, Entebbe, Uganda

[17] Income distribution in OECD countries : evidence from Luxembourg income study / prepared by Anthony B. Atkinson, Lee Rainwater, and Timothy M. Smeeding, Organisation for Economic Co-operation and Development, 1995