Working Paper No. 873

Quality of Match for Statistical Matches Used in the Development of the Levy Institute Measure of Time and Consumption Poverty (LIMTCP) for Ghana and Tanzania

by

Fernando Rios-Avila
Levy Economics Institute of Bard College

September 2016

The Levy Economics Institute Working Paper Collection presents research in progress by Levy Institute scholars and conference participants. The purpose of the series is to disseminate ideas to and elicit comments from academics and professionals.

Levy Economics Institute of Bard College, founded in 1986, is a nonprofit, nonpartisan, independently funded research organization devoted to public service. Through scholarship and economic research it generates viable, effective public policy responses to important economic problems that profoundly affect the quality of life in the United States and abroad.

Levy Economics Institute
P.O. Box 5000
Annandale-on-Hudson, NY 12504-5000
http://www.levyinstitute.org

ISSN 1547-366X

ABSTRACT

This document presents a description of the quality of match of the statistical matches used in the LIMTCP estimates prepared for Ghana and Tanzania. For Ghana, the statistical match combines the Living Standards Survey Round 6 (GLSS6) with the Ghana Time Use Survey (GTUS) 2009, and for Tanzania it combines the Household Budget Survey (THBS) 2012 with the time-use data obtained from the Integrated Labor Survey Module (ILFS) 2006. In both cases, the alignment of the two datasets is examined, after which various aspects of the match quality are described. Despite the differences in the survey years, the quality of match is high and the synthetic dataset appropriate for the time poverty analysis.

Keywords: Statistical Matching; Time Use; Household Production; Poverty; LIMTCP; Ghana; Tanzania

JEL Classifications: C14, C40, D31, J22

I. INTRODUCTION

This paper describes the construction of synthetic datasets created for use in estimation of the Levy Institute Measure of Time and Consumption Poverty (LIMTCP) for Ghana and Tanzania. This work was carried out for a project supported by the Hewlett Foundation to develop and analyze alternatives measures of income/time poverty in developing countries.[1] Construction of LIMTCP estimates requires a variety of information at the household and individual level. In addition to demographic characteristics, the estimation process requires information about income and time use. In order to produce LIMTCP estimates, a synthetic data file is created by statistically matching two source data sets: the base data (recipient data), which contains detailed demographic and income/consumption data for households and individuals; and time-use data (donor data). This creates a unique dataset that would not be available otherwise, from which we can extract patterns of time use for all household members.[2]

For this analysis, the Ghana Living Standards Survey Round 6 (GLSS6) and the Tanzania Household Budget Survey 2012 (THBS) are used as the base/recipient data sets. Both contain good information on demographics and expenditure data that is representative at the national level for all households in their respective countries. Time-use data comes from the Tanzanian Integrated Labor Survey 2006 (ILFS), which contains a specific module collecting time-use data for all family members in Tanzania, and from the Ghana Time Use Survey 2009 (GTUS).

The rest of the paper is organized as follows. Section one describes the data. Section two assesses the alignment of the information between the household surveys and time-use surveys for Ghana and Tanzania. Section three briefly describes the methodology and analyzes the matching quality. Section four concludes.

[1] This work was done under a grant from The William and Flora Hewlett Foundation, "Understanding the Interlocking of Income and Time Deficits for Men and Women in Ghana and Tanzania: Revisiting Poverty Measurement, Rethinking Policy Responses."
[2] See Kum and Masterson (2010) for details of the statistical matching procedure that we use.

1. DATA DESCRIPTION

1.1. Data Sources

For the construction of the LIMTCP estimates we require two different sets of data: household survey data that collects information regarding consumption, demographics, and employment; and time-use data, from which we can estimate the total amount of time households dedicate to household production. A summary of the main characteristics of the data can be found in table 1.

For Ghana, we use the GLSS6, which was conducted in 2012–13. It is a nationally representative dataset that was collected with the purpose of measuring living conditions and wellbeing, and is used to obtain the official poverty profiles in Ghana. The survey is focused on colleting comprehensive data on total household consumption including food, non-food items, and housing costs, adjusting for cost of living differences across the region. A total of 16,772 households and 72,373 persons were included in this survey. As a source for time-use data, we use the GTUS, collected in 2009, the first stand-alone time-use survey conducted in Ghana. The survey had the objective of measuring and analyzing the time spent on all activities of individuals 10 years of age or older over a 24-hour period, with particular emphasis on the gender issues regarding time use. A total of 9,297 persons (of 10 years of age or older) in 4,193 households were interviewed from June to July of 2009.

In the case of Tanzania, the data used to obtain consumption expenditure and employment data comes from the THBS. It is the sixth nationally representative survey conducted in the Tanzanian Mainland (excluding the semiautonomous region of Zanzibar). This survey was conducted with the purpose of collecting household income and expenditure in the country, and implemented from October 2011 to October 2012. Similar to Ghana, their poverty assessment is based on consumption expenditure, adjusting for cost of living differences across geographical areas. The final sample of the survey reached a total of 10,186 households covering 46,593 persons in the country. Finally, the source of time data for Tanzania comes from the ILFS collected in 2006. The ILFS was the first to include an additional module for capturing information on time use (TUS), with the main objective of obtaining data to show the gender disparities in household time use. The TUS module was administered to one in every five households in the ILFS sample, for

seven consecutive days, to all household members ages five years or older. The time survey data was collected for 3,140 households and 10,553 persons ages five years or older.

Table 1. Data Sources, Ghana and Tanzania

Country	Survey Subject	Name	Sample Size
Ghana	Consumption expenditures and employment	Ghana Living Standards Survey (GLSS6) 2012–13	72,373 persons in 16,772 households. There were 52,771 individuals, ages 10 years or older.
	Time use	Ghana Time Use Survey (GTUS) 2009	9,297 persons, ages 10 years or older in 4,193 households. The study used a 24-hour diary, divided into one-hour slots, to record activities. Data was collected from June to July of 2009.
Tanzania	Consumption expenditures and employment	Tanzania Household Budget Survey (THBS) 2011/12	46,593 persons in 10,186 households. There were 39,265 individuals, ages 5 years or older.
	Time use	Integrated Labour Force Survey (ILFS), Time Use Module 2006 (Tanzania Time Use Survey or TUS)	10,553 persons, ages 5 years or older in 3,140 households. Each member of the targeted household was interviewed for seven consecutive days, for their hourly activities during the previous day.

1.2. Methodological Concerns

In order to create the estimates of the time-consumption poverty measures for Ghana and Tanzania, all time thresholds must be constructed at the household level, using total time spent on household production by all its members. For an appropriate identification of the threshold, the reference group consists of households with at least one non-employed adult and income around the official consumption poverty line. For practical purposes, this is identified as households with consumption levels within 75% and 150% of the poverty line (referred to as the poverty band hereafter).

The reference group is divided into subgroups based on the number of children (0, 1, 2, and 3 or more) and number of adults (1, 2, and 3 or more) in order to account for the heterogeneity on the thresholds across different household structures. In principle, they represent the average amount of household production that is required to subsist at the poverty level of income.

In order to appropriately transfer the hours spent by individuals on household production in the reference group as closely as possible, in particular within the reference group and poverty band, we include household-level assets (Tanzania) and household-level income (Ghana). These variables are included as a poverty indicator, or poverty-band indicator, and are not available in the time-use surveys of either country. In addition, the following strata variables are included in the match: indicators for having one or more non-employed adults in the household; the number of children; the number of adults; sex; employment status; geographical area; own labor status; type of household; and relationship to the head of the household.

It should be noted that while the time-use survey in both countries collects relevant information for all members of the household, the data is not matched at the household level, but rather at the individual level. This implies that two individuals from the same family might not be matched to the same household in the recipient data. Nevertheless, during the matching process we include an array of household-level variables that should help improve the quality of the match within the household.

2. DATA ALIGNMENT AND STATISTICS

One of the conditions that needs to be fulfilled before proceeding through a statistical matching process is for the surveys (which are to be statistically merged) to represent the same population, with approximately similar characteristics across their weighted samples. If this were not to be true, the donor or recipient data would need to be reweighted for the synthetic data to be appropriate for inferences. In this section we present a set of statistics to assess the comparability of the household surveys and time-use data for Ghana and Tanzania.

2.1. Data Alignment

Tables 2 and 3 compare the distribution of individuals across selected characteristics for both Ghana and Tanzania, respectively, including the strata variables. Despite the fact that the surveys were collected in different years (with a three-year gap for Ghana's surveys and a six-year gap for Tanzania's), we can still expect them to be relatively well aligned, as most of the variables used should reflect structural characteristics that are rather stable across time.

When looking at Ghana (table 2), we see an oversampling of households without children in the GTUS, although that is only observed for households with only one adult (8 percentage point difference). On the other extreme of the spectrum, the GLSS6 indicates there are more people living in large households (5 percentage points more). In terms of share of people by age, education, and sex, the surveys seem to be well balanced. Regarding labor status, the GLSS6 indicates there is a larger share of the population who are classified as employed (5.2 percentage points) and that there are fewer unemployed workers. Despite this difference, however, there is good alignment regarding total number of people living in households where there is at least one non-working adult member. In terms of household structure and member's relationship to the head, the GLSS6 and GTUS show some imbalances, with a larger share of single-headed households shown in the time-use survey (5 percentage points). While the share of households with a domestic servant living in the house is small in Ghana, the time-use data indicates a proportion more than twice as large as the GLSS6 might suggest.

For Tanzania (table 3) we see that there are some differences with respect to the household structure and number of children in the household. The most notable is that there are more people living in households with no children according to the ILFS data, while there is an almost 7 percentage point difference in the number of people living in the large households (more than three adults and more than three children) according to the THBS. Another important difference is that the THBS indicates that there is at least a 7 percentage point difference in the number of people living in households where all adults are working, which reflects the higher rates of people not participating in the labor force in the THBS survey. It is also noticeable that the THBS seems to have oversampled households with more children, while undersampling single-headed households, which can also be observed by the differences in the balance of share of

people by household type and relationship to the head. In terms of educational attainment, the data in the THBS indicates a higher level of education for the population as a whole, which could be partially explained by a secular improvement in human capital accumulation in the country, with the proportion of people with secondary education almost twice as large in the THBS compare to the ILFS data.

Table 2. Alignment for Ghana: GLSS6 and GTUS

	GTUS	GLSS6		GTUS	GLSS6
Adult X, Child			**Sex**		
1 Adult, 0 Children	28.4	20.2	Male	48.1	48.1
1 Adult, 1 Children	5.5	4.5	Female	51.9	51.9
1 Adult, 2 Children	4.8	3.3			
1 Adult, 3+ Children	3.8	3.4	**Non-working adults**		
2 Adults, 0 Children	7.7	8.4	All Working	69.3	67.5
2 Adults, 1 Children	9.8	9.1	Some Non-working	30.8	32.5
2 Adults, 2 Children	9.3	9.4			
2 Adults, 3+ Children	14.2	15.3	**Labor Force Participation**		
3+ Adults, 0 Children	2.7	4.3	Employed	58.1	63.8
3+ Adults, 1 Children	3.3	4.7	Unemployed	4.9	2.5
3+ Adults, 2 Children	3.3	4.9	Not in LF	37.0	33.7
3+ Adults, 3+ Children	7.4	12.6			
Age Group			**Rel. to the Household Head**		
0–10	31.5	29.1	Head	29.7	24.8
11–20	19.4	23.1	Spouse	14.3	13.7
21–30	15.2	15.5	Children	44.2	46.9
31–40	12.8	11.9	Other	11.1	13.7
41–50	9.0	8.7	Domestic	0.8	0.8
51–60	5.8	5.8			
61–70	3.5	3.2	**HH Type**		
71+	2.9	2.9	Single	53.7	47.8
Education Level			Married	44.6	49.3
Never Attended/Other	20.04	21.03	Polygamous	1.7	2.9
Primary	22.33	26.5			
Middle	35.73	32.94	**HH with Domestic Servants**		
Sec. or Above	21.9	19.53	No Domestic Servant	98.44	99.35
Geographical Area			With Domestic Servant	1.56	0.65
Urban	53.69	55.39			
Rural	46.31	44.61			

Table 3. Alignment for Tanzania: THBS and ILFS (shares)

	ILFS	THBS		ILFS	THBS
Adult X Child			**Sex**		
1 Adult, 0 Children	12.13	9.44	Male	47.9	48.5
1 Adult, 1 Children	4.87	3.32	Female	52.1	51.5
1 Adult, 2 Children	3.25	2.9			
1 Adults, 3+ Children	4.04	3.06	**Non-Working Adult**		
2 Adults, 0 Children	8.68	6.79	All Adults Working	82.59	75.1
2 Adults, 1 Children	11.28	10.32	1+ Non-working Adults	17.41	24.9
2 Adults, 2 Children	12.13	10.97			
2 Adults, 3+ Children	19.66	20.92	**Labor Force Participation**		
3+ Adults, 0 Children	4.01	3.42	Employed	70.02	69.28
3+ Adults, 1 Children	4.14	4.77	Unemployed	2.8	1.6
3+ Adults, 2 Children	4.03	5.37	Not in LF	27.18	29.12
3+ Adults, 3+ Children	11.78	18.72			
Age Group			**Rel. to the Household Head**		
0–10	18.04	19.73	Head	29.13	24.21
11–20	25.97	27.2	Spouse	18.95	16.91
21–30	18.91	18.11	Children	37.19	41.59
31–40	14.03	13.92	Other	14.25	16.48
41–50	9.48	8.93	Domestic	0.5	0.82
51–60	6.3	5.65			
61–70	4.19	3.48	**Household Type**		
71–80	2.11	2	Single	36.51	32.19
81+	0.97	0.97	Couple	61.85	65.96
Education Distribution by survey			Polygamous	1.65	1.85
Never Attended	24.68	20.08			
Primary Not Complete	36.17	32.83	**HH With Domestic Servants**		
Primary Complete	32.95	34.34	No Domestic Servant	95.81	94.83
Secondary Above	6.2	12.76	With Domestic Servant	4.19	5.17
Geographical Area					
Capital	10.05	10.7			
Urban	16.31	19.47			
Rural	73.64	69.83			

The majority of the statistics presented here suggest that there is reasonably good alignment between the household surveys and the time-use data in both countries. There are, however, some consistent misalignments that need to be addressed to reduce the possible bias they could generate. To alleviate these imbalances, the data survey weights are adjusted so that the time-use survey has the same structure as the household budget survey based on the household composition (i.e., a combination between the number of adults and the number of children) and the sex of the person before the statistical match is applied. Nevertheless, the quality match statistics shown in the next section compare the imputed and real distribution of time use based on the original survey weights.

3. MATCH QUALITY

3.1. Methodology

Statistical matching (also known as data fusion) is a widely used technique in empirical studies and has been applied in cases when no single survey contains all the relevant information needed for drawing important inferences. There are numerous empirical works in the economic field that have applied this strategy (see, for example, Rässler [2002] and, more recently, D'Orazio, Di Zio, and Scanu [2006]).

This method, which is similar to single imputation methods, consists of combining the information of two separate and independent surveys into a single combined dataset from which statistical inferences can be obtained. It enables the combination of the datasets using common information between both surveys, while trying to preserve the distributional characteristics of the combined information under the assumption that both surveys represent the same population.

The algorithms that can be used to perform statistical matching can broadly be classified into two groups. The first one is known as "unconstrained statistical matching." This strategy frequently uses some type of distance criterion (propensity score matching, for example) so that the best possible candidate (based on observable characteristics) is chosen (often with replacement) from the donor file to be matched with the corresponding recipient observation.

The second group is known as "constrained statistical matching." In this case, the strategy imposes the restriction that all observations, specifically their weighted representation from both the donor and recipient surveys, need to be used in the final match. This strategy often relies on a rank imputation, using broad strata variables to avoid undesirable matches.[3] This paper uses the methodology proposed in Kum and Masterson (2010), which has been used in the estimation of the Levy Institute Measure of Economic Well-Being (LIMEW) (Wolff and Zacharias 2003) and Levy Institute Measure of Time and Consumption Poverty (LIMTCP) (Zacharias, Masterson, and Memis 2014).

[3] The hot deck matching uses ranked information based on some auxiliary information such as the propensity score. For further details on the matching procedure, see Kum and Masterson (2010).

3.2. Matching Rounds

We now turn our attention to the match process and results. The first thing to look at is the distribution of matched records by matching round. While one would prefer to have a larger share of the observations be matched during the first steps of the matching algorithm, thus ensuring a higher quality match, the rate at which observations are matched can vary based on the restrictions on strata variables that can be imposed. Figures 1 and 2 present the share of observations in the household budgets that are matched during each round for Ghana and Tanzania, respectively.

For Tanzania, a bulk of the matches, 58%, occur in the first round, while for Ghana just below 4% are matched in the first round. This is lower than in other time-use matches, particularly for Ghana (see, for example, Masterson [2010]) due to the higher than usual number of variables used in this first round of the match.[4] For Tanzania, over 80% of the observations are matched by round 4, with small match changes thereafter. After 21 rounds, 0.7% of the recipient data in Tanzania receives no match at all. In the case of Ghana, the bulk of the matching occurs between rounds 7 to 18, with 80% of the data being matched after this step.[5] At the end of the matching process, only one observation from the Ghana household survey was left unmatched. It should be noted that approximately 1.4% (Tanzania) and 2.4% (Ghana) of the time-use data was not used for the match.

[4] In a typical time-use match (as in Masterson [2010]), only five variables are used, yielding a total of 32 matching cells. In Ghana's case, seven strata variables are used, yielding a total of 781 of matching cells in the first round. For a similar pattern of match, see Rios-Avila (2015).

[5] When comparing the performance of the statistical matching process across both countries, it should be noticed that we are able to include an income variable index for Ghana, which increased the complexity of the match, explaining the observed differences in the distribution of matched records.

Figure 1. Distribution of Matched Records by Matching Round, Ghana

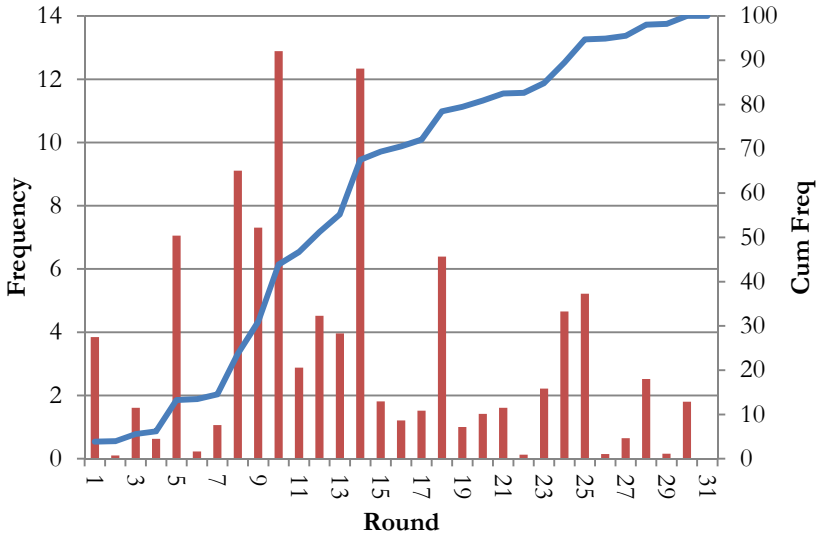
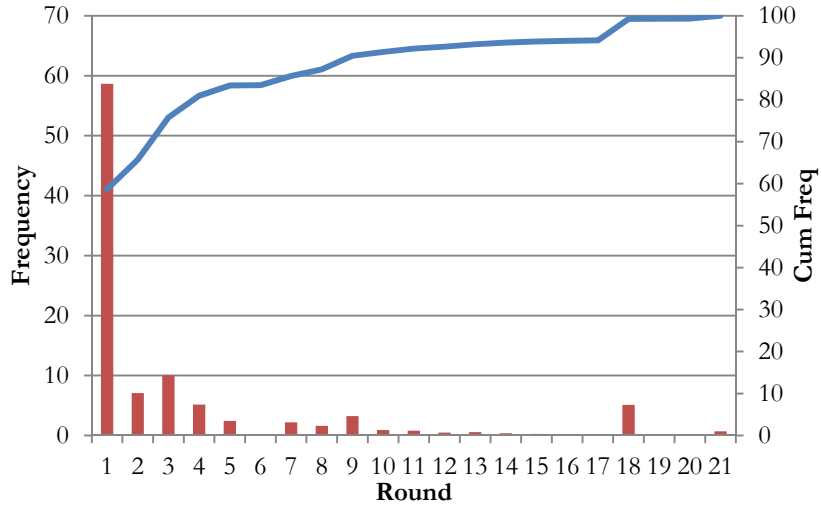


Figure 2. Distribution of Matched Records by Matching Round, Tanzania



3.3. Match Quality Assessment

We now turn our attention to the assessment of the match quality for Ghana and Tanzania. According to Rässler (2002, 2004), the quality of the statistical matching can be assessed based on the ability of the match to preserve the true individual values of the distribution (strongest test), capacity to preserve the joint distribution of the transferred data, preserve the correlation of the data, or preserve their marginal distributions (weakest test). Since the true values of the transferred data, as well as the joint distribution or correlations, are unknown, we need to apply statistical matching, and we will assess the quality of the match based on comparisons of the

marginal distribution of the transferred data (time use) across various selected household characteristics.

While there are different strategies that have been developed to assess the quality of the transferred data, including the comparison of the coefficients of potential explanatory econometric models (see Rios-Avila [2015] for an example of this alternative), in this paper we will use a different method. This method consists of analyzing the ratio of the average (or some selected percentiles) value of the transferred variables over the true averages in the original sample. In this case, we report the ratios of household production components across many selected characteristics and strata variables.[6]

Table 4. Distribution of Weekly Hours of Household Production in TUS and Matched File

Ghana	Gini	p90/p10	p90/p50	p50/p10	p75/p25	p75/p50	p50/p25
Match	0.588		4.24		42.00	2.44	17.20
Time Use	0.591		4.29		105.00	2.47	42.50
Tanzania	**Gini**	**p90/p10**	**p90/p50**	**p50/p10**	**p75/p25**	**p75/p50**	**p50/p25**
Match	0.508	87.00	3.35	26.00	6.65	2.22	3.00
Time Use	0.503	88.67	3.28	27.00	6.41	2.14	3.00

Table 4 provides a comparison of the distribution of weekly hours of household production in the donor and matched file for the overall distribution for both Ghana and Tanzania. For Ghana, the tenth percentile is zero, so the p50/p10 and p90/p10 ratios are undefined. The remaining percentile ratios are all relatively close, with the exception of the ratio in the p75/p25 and p50/p25 percentile for Ghana. The main reason for this is that households at the 25th percentile spend very little time on household production (about 0.6 hours according to the household survey and 0.2 hours in the time-use data), which makes the ratio very sensitive to small differences in that percentile. For Tanzania, all ratios are very well aligned. The Gini coefficient is extremely close, only 0.005 (Tanzania) and 0.003 (Ghana). The close balance of the overall distribution is expected, as the matching process guarantees an almost-perfect transference of the overall distribution from the time use to the household survey data.

[6] Ratios around 100% are considered signals for good quality of the match. As a rule of thumb, we use a confidence interval from 80–120% for such assessment.

Table 5. Comparison of Mean Time-Use Variables and Matched Data

Ghana	Core	Procurement	Care	Wood Collection	Water Collection	Commuting
Match	10.68	1.06	3.47	0.41	0.54	1.45
Time Use	10.66	1.04	3.54	0.41	0.46	1.57
Ratio	100.1%	102.2%	97.9%	99.4%	117.9%	92.1%
Tanzania	**Core**	**Procurement**	**Care**	**Wood Collection**	**Water Collection**	**Commuting**
Match	11.18	1.55	2.84	0.73	1.80	1.06
Time Use	11.63	1.60	2.79	0.72	1.81	1.15
Ratio	96.1%	96.7%	101.8%	100.6%	99.7%	91.9%

Table 5 breaks down household production into the three categories, namely care (child care, elder care, etc.), procurement (shopping, etc.), and core (cooking, cleaning, laundry, etc.), and compares donor-recipient ratio averages for both countries. We also include some additional information on time the households spend collecting wood and water for personal use, as well time spent commuting. We see that for all the time-use aggregates, the differences in the averages between the synthetic and original file mean are small, with the largest proportional difference observed for Tanzania (3.9% in core and 3.3% in procurement). We do observe some larger imbalances when looking at specific time-use categories, including water collection (Ghana) and commuting (Ghana and Tanzania), although in both cases the absolute differences are small (less than 0.1 hours a week) and shouldn't be considered a problem in the match.

In figures 3 and 4, we present boxplot representations of the distribution of time use on household production by the main strata variables, number of adults, and number of children. In the context of the LIMPCP, these are the most important variables across which we require a good quality of the match, since time poverty thresholds are constructed using these demographic groups. A visual inspection of these data also suggests the quality of the match data is high, as the median and the 25th and 75th percentile are very close together for both countries, with few observable misalignments around the edge of the distribution.

Figure 3. Household Production by Reference Groups, Matched vs. Time Use, Ghana

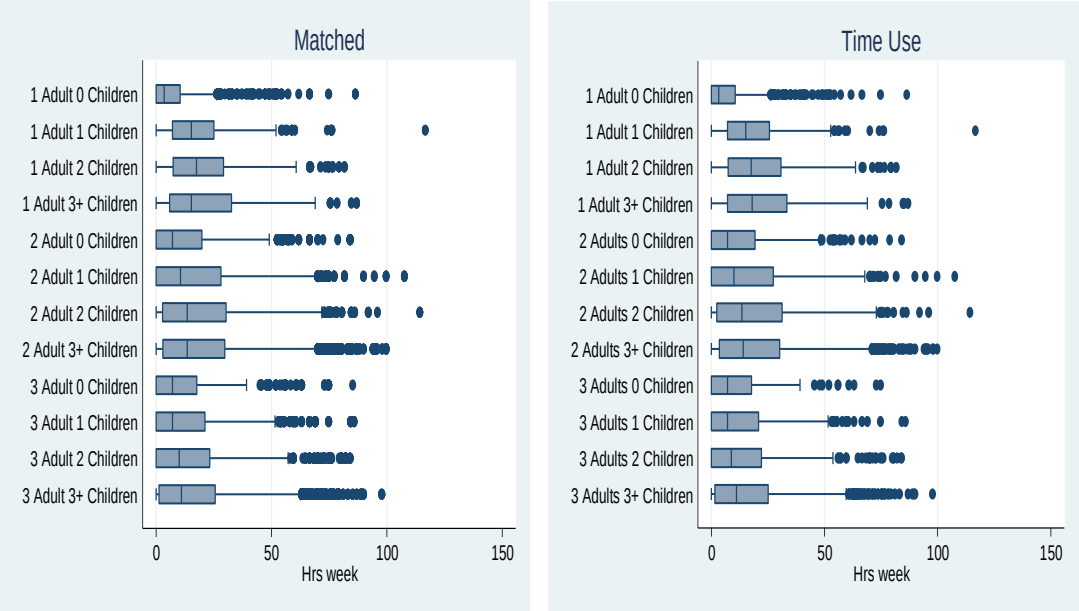
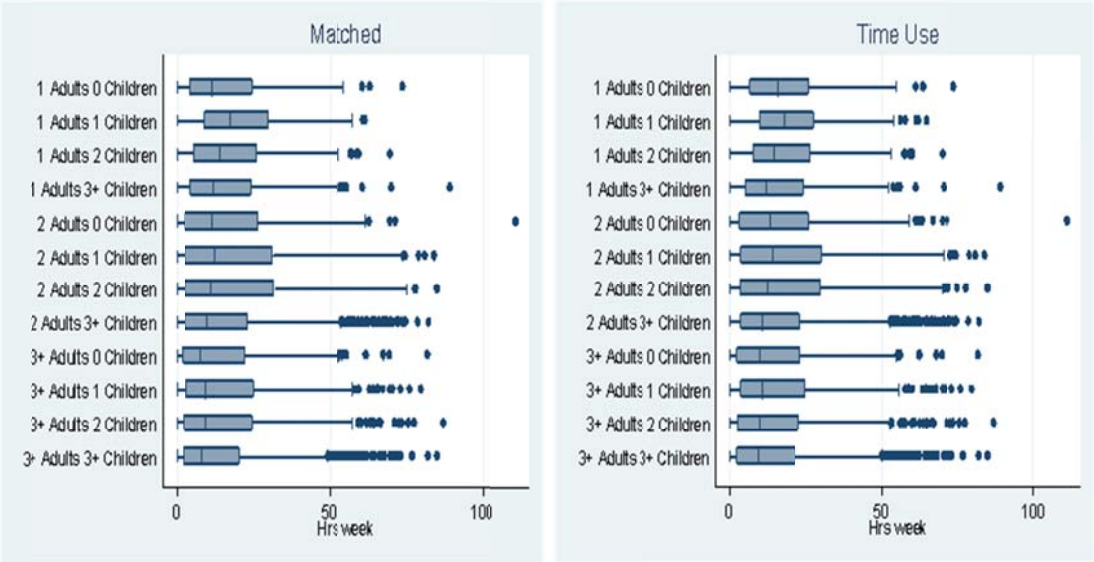


Figure 4. Household Production by Reference Groups, Matched vs. Time Use, Tanzania



An alternative analysis to figures 3 and 4 is to estimate the average (median) household production ratios conditional on a few of important strata variables. In tables 6 and 7, we present these ratios for Ghana and Tanzania, respectively, as a function of the number of children, the number of adults, household employment status, individual employment status, sex, and geographical area. We can see that all the ratios related to these selected variables fall within a

very narrow band of under 10% for the average ratios. The largest differences can be observed among people living in the capital of Ghana, with a relative gap of 7.3% (0.98 hours), and among non-employed workers in Tanzania, with a relative gap of 6.1% (0.74 hours). Looking at the median ratios, we do observe a couple of abnormally large ratios, such as a 14% gap for employed people (1.2 hour gap), a 100% gap for men (0.6 hour gap) in Ghana, and a 9.6% difference for employed people in Tanzania (1.3 hours).

Table 6. Mean and Median Weekly Hours of Household Production by Selected Strata Variables, Ghana

Average						
Number of Children	Matched Data	Time-Use Data	Ratio		Matched Data	Time-Use Data
0 Children	8.93	8.77	101.8%			
1 Child	14.22	14.04	101.2%	1/0	1.59	1.60
2 Children	16.26	16.36	99.4%	2/0	1.82	1.86
3 Children	16.43	16.21	101.4%	3/0	1.84	1.85
4 + Children	15.71	15.65	100.4%	4+/0	1.76	1.78
Number of Adults						
1 Adult	12.50	12.57	99.5%			
2 Adults	16.51	16.39	100.7%	2/1	1.32	1.30
3 or More Adults	13.37	13.24	101.0%	3+/1	1.07	1.05
Non-employed Adult in Household (y/n)						
No	14.73	14.81	99.5%	yes/no	0.96	0.93
Yes	14.10	13.82	102.0%			
Employed (y/n)						
No	14.34	14.46	99.1%	yes/no	1.03	0.98
Yes	14.72	14.19	103.7%			
Sex						
Male	5.26	4.98	105.6%	Fem/Male	4.30	4.55
Female	22.60	22.66	99.7%			
Rural/Urban						
Capital	13.59	12.67	107.3%	Capital/urb	0.92	0.85
Urban	14.74	14.83	99.4%	Rur/Urb	0.99	1.00
Rural	14.57	14.80	98.5%			
Median						
Number of Children	Matched Data	Time-Use Data			Matched Data	Time-Use Data
0 Children	3.50	3.50	100.0%			
1 Child	8.17	8.17	100.0%	1/0	2.33	2.33
2 Children	10.50	10.50	100.0%	2/0	3.00	3.00
3 Children	10.50	9.92	105.9%	3/0	3.00	2.83
4 + Children	9.92	9.92	100.0%	4+/0	2.83	2.83
Number of Adults						
1 Adult	7.00	7.00	100.0%			
2 Adults	10.50	9.92	105.9%	2/1	1.50	1.42
3 or More Adults	7.58	7.00	108.3%	3+/1	1.08	1.00
Non-employed Adult in Household (y/n)						
No	8.75	9.33	93.8%	yes/no	0.87	0.75
Yes	7.58	7.00	108.3%			
Employed (y/n)						
No	8.05	8.17	98.6%	yes/no	1.16	1.00
Yes	9.33	8.17	114.3%			
Sex						
Male	1.17	0.58	200.0%	Fem/Male	16.00	32.00
Female	18.67	18.67	100.0%			
Rural/Urban						
Capital	7.00	6.42	109.1%	Capital/urb	0.86	0.76
Urban	8.17	8.46	96.6%	Rur/Urb	1.14	1.10
Rural	9.33	9.33	100.0%			

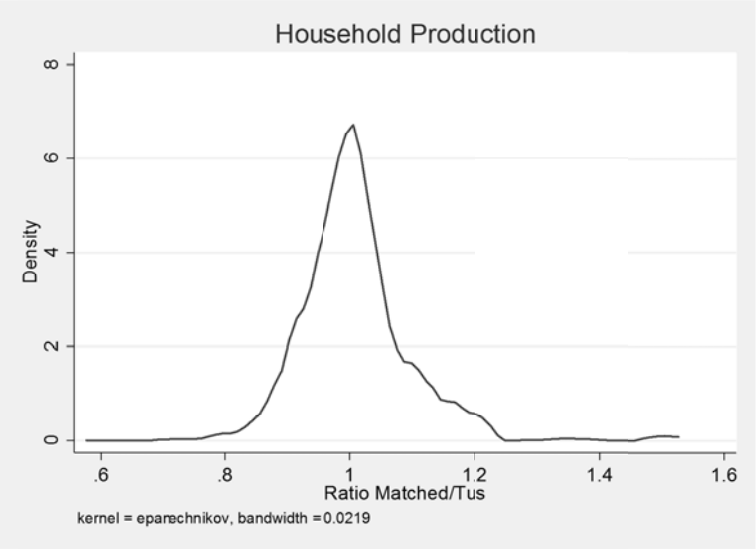
Table 7. Mean and Median Weekly Hours of Household Production by Selected Strata Variables, Tanzania

Average						
	Matched	Time-Use			Matched	Time-Use
Number of Children	Data	Data	Ratio		Data	Data
0 Children	16.02	16.24	98.6%			
1 Child	17.99	18.38	97.8%	1/0	1.12	1.13
2 Children	17.03	17.31	98.4%	2/0	1.06	1.07
3 or More Children	14.61	14.83	98.5%	3+/0	0.91	0.91
Number of Adults						
1 Adult	17.15	17.52	97.9%			
2 Adults	16.56	16.78	98.7%	2/1	0.97	0.96
3 or More Adults	14.46	14.54	99.5%	3+/1	0.84	0.83
Non-employed Adult in Household (y/n)						
No	15.76	16.25	97.0%	yes/no	0.96	0.93
Yes	15.08	15.14	99.6%			
Employed (y/n)						
No	12.80	12.06	106.1%	yes/no	1.31	1.47
Yes	16.80	17.71	94.8%			
Sex						
Male	7.31	7.38	99.0%	Fem/Male	3.20	3.24
Female	23.40	23.95	97.7%			
Rural/Urban						
Capital	15.51	16.17	95.9%	Capital/urb	0.89	0.90
Urban	17.36	17.92	96.9%	Rur/Urb	0.87	0.87
Rural	15.08	15.58	96.8%			
Median						
	Matched	Time-Use			Matched	Time-Use
Number of Children	Data	Data	Ratio		Data	Data
0 Children	12.00	12.50	96.0%			
1 Child	12.67	13.50	93.8%	1/0	1.06	1.08
2 Children	11.08	11.50	96.4%	2/0	0.92	0.92
3 or More Children	9.83	10.00	98.3%	3+/0	0.82	0.80
Number of Adults						
1 Adult	14.00	14.47	96.8%			
2 Adults	11.00	11.28	97.5%	2/1	0.79	0.78
3 or More Adults	9.50	9.38	101.2%	3+/1	0.68	0.65
Non-employed Adult in Household (y/n)						
No	11.00	11.50	95.7%	yes/no	0.86	0.78
Yes	9.50	9.00	105.6%			
Employed (y/n)						
No	8.00	7.58	105.5%	yes/no	1.49	1.74
Yes	11.90	13.17	90.4%			
Sex						
Male	4.50	4.67	96.4%	Fem/Male	4.67	4.65
Female	21.00	21.70	96.8%			
Rural/Urban						
Capital	8.00	7.58	105.5%	Capital/urb	0.73	0.67
Urban	11.02	11.38	96.8%	Rur/Urb	0.95	0.98
Rural	10.50	11.20	93.8%			

While the results in tables 6 and 7 provide a reasonable overview of the quality of the transferred time-use data, it is still rather restrictive, as it only shows results for a select group of variables. In an attempt to provide a more comprehensive overview of the marginal distribution across all the categorical variables involved in the matching process, including a few variable

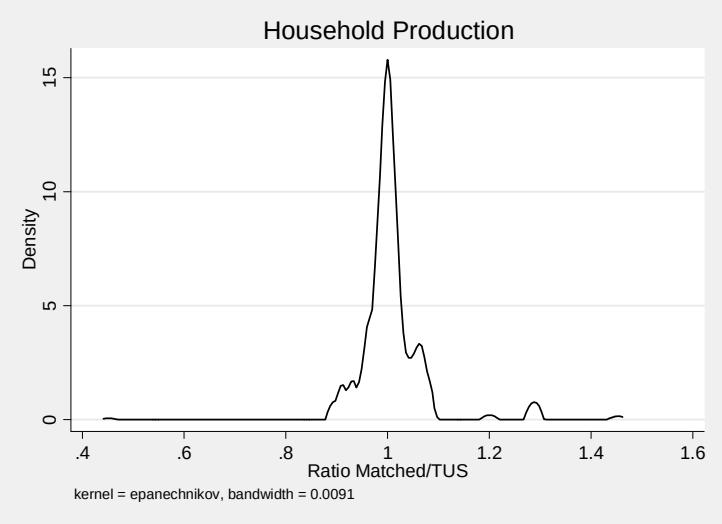
combinations, we provide a kernel density of all ratios, weighted by the number of observations used for the ratio in each category. Thus, ratios that represent a more comprehensive group, say men or women, will have more weight on the kernel density compared to the ratio of young unemployed women living in rural areas. This information is presented in figures 5 and 6.

Figure 5. Density of Ratios of Mean Household Production (matched/time use), Ghana



Note: Figure uses kernel density of the ratio of hours of household production, based on different variables used in the match. Each ratio is weighted by the number of people in that particular group.

Figure 6. Density of Ratios of Mean Household Production (matched/time use), Tanzania



Note: Figure uses kernel density of the ratio of hours of household production, based on different variables used in the match. Each ratio is weighted by the number of people in that particular group.

Based on figures 5 and 6, the examination of the quality of the match within population subgroups shows generally good results. In both countries, the largest share of the distribution falls within the 10% difference between the matched and donor data, and almost all the ratios fall within the 20% difference. While there are a few observations that fall beyond this interval for both countries—observed in the long tails of the distribution—they represent very small segments of the population or often involve gaps that are small in magnitude (less than 1 hour per week). It should be expected that such a difference would not affect the conclusions of the matching results.

4. CONCLUSIONS

This paper presents the application and quality assessment of the statistical matching algorithm used to combine the household survey data and time-use data for Ghana and Tanzania as part of the project to obtain comprehensive estimates on time and income poverty. Overall, for both Ghana and Tanzania, the household survey data and time-use data are well aligned, with some imbalances with respect to household family size and presence of children in the household. These imbalances are corrected before the implementation of the statistical matching using an inverse probability weighting strategy.

Based on the statistics presented here, the matching quality is good, showing strong balance across different household characteristics. There are, however, a few large imbalances that are isolated in small groups. Despites its limitations for some small groups, in general the statistical matching procedure does a good job in transferring the distribution of hours of household production for both countries.

REFERENCES

D'Orazio, M., M. Di Zio, and M. Scanu. 2006. *Statistical Matching: Theory and Practice*. Chichester, UK: Wiley.

Kum, H., and T.N. Masterson. 2010. "Statistical matching using propensity scores: Theory and application to the analysis of the distribution of income and wealth." *Journal of Economic and Social Measurement* 35(3): 177–96.

Masterson, T. 2010. "Quality of Match for Statistical Matches Used in the 1992 and 2007 LIMEW Estimates for the United States." Levy Institute Working Paper 618. Annandale-on-Hudson, NY: Levy Economics Institute of Bard College.

Rässler, S. 2002. *Statistical matching: A frequentist theory, practical applications, and alternative Bayesian approaches*. New York: Springer.

Rässler, S. 2004. "Data Fusion: Identification Problems, Validity, and Multiple Imputation." *Austrian Journal of Statistics* 33(1, 2): 153–71.

Rios-Avila, F. 2015. "Quality of Match for Statistical Matches Using the Consumer Expenditure Survey 2011 and Annual Social Economic Supplement 2011." Levy Institute Working Paper 830. Annandale-on-Hudson, NY: Levy Economics Institute of Bard College.

Wolff, E.N., and A. Zacharias. 2003. "The Levy Institute Measure of Economic Well-Being." Levy Institute Working Paper 372. Annandale-on-Hudson, NY: Levy Economics Institute of Bard College.

Zacharias, A., T. Masterson, and E. Memis. 2014. "Time Deficits and Poverty: The Levy Institute Measure of Time and Consumption Poverty for Turkey." Levy Institute Research Project Report, May. Annandale-on-Hudson, NY: Levy Economics Institute of Bard College.