Quality of Statistical Match and Simulations Used in the Estimation of the Levy Institute Measure of Time and Consumption Poverty (LIMTCP) for Turkey in 2006

by

Thomas Masterson
Levy Economics Institute of Bard College

July 2013

ABSTRACT

The quality of match of the statistical match used in the Levy Institute Measure of Time and Consumption Poverty (LIMTCP) estimates for Turkey in 2006 is described. The match combines the 2006 Zaman Kullanim Anketi (ZKA 2006) with the 2006 Hanehalki Bütçe Anketi (HBA 2006). These are the national time-use survey and household income and expenditure surveys, respectively. The alignment of the two datasets is examined, after which various aspects of the match quality are detailed. The match is of high quality, given the nature of the source datasets.

The quality of the simulation of employment gains for Turkey in 2006 is then described. All eligible adults not working for pay, as employers, or as unpaid household workers were assigned jobs. In all households that included job recipients, the time spent on household production was imputed for everyone included in the time-use survey. Household consumption was then assigned to each household in the simulation containing a job recipient. The recipient group was compared to the donor group, both in terms of demographic similarity and in terms of the imputed usual hours, earnings, and household production generated in the simulation. In both cases, the simulations were of reasonable quality, given the nature of the challenges in assessing their quality.

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 Turkey in 2006. This work was carried out for a project supported by the United Nations Development Programme.[1] Construction of LIMTCP estimates requires a variety of information for households. In addition to basic demographics, the estimation process requires information about income and time use. No single data set has all the required data for Turkey. Thus, in order to produce LIMTIP estimates, a synthetic data file is created by statistically matching two source data sets.[2] We use the Hanehalki Bütçe Anketi (HBA 2006, the household income and expenditure survey) carried out by the Türkiye İstatistik Kurumu (TÜIK, the Turkish statistical institute) as the base data set, since it contains good information on demographics, income, transfers and taxes for a representative sample of households in Turkey. Time use data comes from TÜIK's time use survey Zaman Kullanim Anketi (ZKA 2006), which is also nationally representative. With the resulting file we estimate time and consumption poverty.

In order to assess the possible impact of income-poverty reduction strategies founded upon expanding employment on time and consumption poverty, it is necessary to impute the impact of those strategies on the income, time allocation, and consumption expenditures of households. We draw on and extend our work simulating the results of the American Recovery and Reinvestment Act (Zacharias, Masterson and Kim 2009) and previous LIMTIP employment simulations (Masterson 2012). In this case, we assume that some unspecified way is found to employ those adults in households underneath our adjusted consumption poverty line who are not employed. We then assess the impact this change has on time and consumption poverty.

This paper is organized as follows. The source datasets are described and their demographic characteristics are compared. Then the quality of the match is reviewed including diagnostics about the match itself. Next we describe the methodology involved in the imputation of occupation and industry, hours of employment and earnings, household income, household production hours, and consumption expenditures. Finally, we assess the results of the simulation.

---

[1] The project, titled "Why Time Deficits Matter: Implications for Understanding and Combatting Poverty", is directed by Ajit Zacharias and Rania Antonopoulos.
[2] See Kum and Masterson (2010) for details of the statistical matching procedure that we use.

STATISTICAL MATCHING

Data and Alignment

The source data sets for the time use match for the LIMTIP estimates for Turkey are the 2006 HBA and the 2006 ZKA. We use individual records from the 2006 HBA file, excluding those living in group quarters or in the Armed Forces. Since the ZKA covers individuals aged 15 to 99 years old, we discard younger and older individuals from the HBA file. This leaves 24,867 records, which represents 51,674,609 individuals when weighted. In the ZKA, we have 10,893 individual records, representing 47,443,177 individuals when weighted.

In order to create the estimates of the time-income poverty measure, we had to construct thresholds for the time spent on household production. The thresholds are defined for the household. The reference group in constructing the thresholds consists of households with at least one nonemployed adult and income around the official income-poverty line. We divided the reference group into 12 subgroups based on the number of children (0, 1, 2 and 3 or more) and the number of adults (1, 2 and 3 or more) for calculating the thresholds. The thresholds are simply the average values of the time spent on household production by households, differentiated by the number of adults and children. In principle, they represent the average amount of household production that is required to subsist at the poverty-level of income.

For practical purposes, we defined the reference group as households with household incomes between 75 percent and 150 percent of the poverty line (this range is referred to as the poverty band hereafter), and with at least one non-employed adult. In order to transfer the hours spent by individuals on household production in the reference group as closely as possible, we used the following strata variables in the match: indicators for being within the poverty band, for having one or more non-employed adults in the household, the number of children, the number of adults, sex, employment status, and household income category. Table 1 compares the distribution of individuals by these variables in the two data sets. Since both surveys were carried out over roughly the same time period, we can expect them to be well aligned. We see that there are 3 percent fewer individuals in households without children in the HBA than in the ZKA, while individuals in two and three-or-more children households make up a greater share of those in the income and expenditure survey. Individuals in one-adult and two-adult households are more common in the ZKA and those in three-or-more-adult households are more common in the HBA. The ratio of individuals in households with at least one non-employed

adult differs by less than half of one percent between the two surveys, while the ratio of individuals in households within the poverty band is 2 percent higher in the ZKA. The distribution by household income is noticeably skewed to the lower end of the distribution in the ZKA compared to the HBA. This is due to the poor quality of the household income question and data in the time use survey. The nonemployed are slightly under-represented in the ZKA relative to the HBA (1.4 percent). The distribution of individuals by sex is close in the two surveys, with females slightly less common (1.5 percent) in the ZKA than in the HBA. So, as expected, we have a very close alignment between the two surveys along six of the seven strata variables.

Quality Assessment

Turning to the results of the match, we first look to the distribution of matched records by matching round in Table 2. The bulk of the matches, 65.2 percent, occur in the first round. This is lower than in other time use matches (see, for example, Masterson 2010), due to the higher than usual number of strata variables used in this match.[3] The rest of the records are matched over an additional 51 rounds, with 0.5 percent receiving no match at all. Table 3 provides a comparison of the distribution of weekly hours of household production in the ZKA and the matched file. The tenth percentile is zero, so those ratios are undefined. The remaining percentile ratios are all relatively close, with the ratio of the median to the 25th percentile being exactly equivalent. The Gini coefficient is extremely close, 0.5519 in the matched file, compared to 0.5521 in the ZKA. Table 4 breaks down the mean and median of the three categories of household production and the total in the matched file and the ZKA.[4] We can see that for all four variables the difference in the matched and the source file's mean is very small, with the largest proportional difference in procurement, which is 0.6 percent (or 36 seconds) higher in the matched file than in the ZKA. Median core and total household production is exactly equal in the matched file.

Examination of the quality of the match within population sub-groups shows generally good results. Figure 1 displays ratios of mean weekly hours of household production between the matched file and the ZKA for the seven strata variables. For almost all the categories, the

---

[3] In a typical time use match (as in Masterson 2010), five variables are used, yielding a total of 32 matching cells. In this match, using seven strata variables, the number of matching cells in the first round was 781.
[4] The three categories are care (child care, elder care, etc.), procurement (shopping, etc.), and core (cooking, cleaning, laundry, etc.).

average weekly hours in the matched file are within 5 percent of the ZKA. The exceptions are for females, who have 6.2 percent higher weekly hours in the matched file, while males have 3.3 percent lower weekly hours. The largest differences are for the top two household income categories, at 8.1 percent and 8.7 percent. This is not surprising, given the relationship between household income and household production weekly hours (more on that below), and the misalignment of the two surveys along this axis.

Table 5 has the actual numbers, and we can see that these large percentage differences represent relatively small differences in hours per week. For the largest percentage gap, in the highest income category, we see that the actual amount of time difference is one and a half hours per week. In the case of sex, males have 25 more minutes per week on average in the matched file, while females have 72 minutes more. Notice that the ratios by category are well reproduced in the matched file. The largest deviation is by sex, as we would expect given the differences in the averages for females and males. The extent to which the match file reproduces the distribution of weekly hours of household production within reference groups is demonstrated in Figure 2 and Table 6.[5] We can see very little difference between the matched file and the ZKA in the distributions in Figure 2, although the average values of weekly household production hours in the matched file range from one percent lower to almost sixteen percent higher than in the ZKA. Thus the distribution of household production is well preserved in the matching process, even at this level of detail.

Overall, the quality of the match is very good. It has its limitations, especially in terms of household income. But the overall distribution is transferred with reasonable accuracy, and the distributions within even small sub-groups, such as one adult with two children, is transferred with good precision.

LABOR MARKET SIMULATION

Data and Methods

The purpose of the simulation is to assess the first order impacts of policies aimed at alleviating consumption poverty in Turkey via jobs policies, for example an employer of last resort (ELR) policy. As such, the simulation is a three-step procedure. The first step is imputing the earnings and the hours of employment of those to be assigned jobs, and adjusting the household income

---

[5] For the sake of clarity of the plot, only the number of children and number of adults is used.

of households with members who have been assigned jobs. The second step is to impute the new hours of household production of individuals in households affected by job assignments. The third step is to impute the new level of consumption for the households with job recipients. With these three steps completed, we can estimate the impact of a given policy on time and consumption poverty, both overall and on individual households. We first discuss the policy scenario, then the steps involved in constructing the estimated outcome of the policy.

Policy Scenario

A very simplified job assignment scenario is envisioned in the LIMTCP Turkey project: that all eligible adults[6] in households below the adjusted consumption poverty line that are not working receive paid (either formal or informal) employment.[7] The donor pool contains all those currently working for pay. After eligible adults are assigned a job, with hours and earnings, the household income of households with eligible adult(s) is recalculated by adding the imputed amount of household earnings to the previous amount of household income. We assume that none of the other components (i.e. other than earnings) of household income undergo any change (i.e., we incorporate the maximum income effect of additional employment in our simulation). This assumption is, obviously, unrealistic for households that receive means-tested income transfers or receive income transfers that depend on employment status. However the HBA 2006 does not provide detailed data on transfers, just the total amounts received in cash or in kind. Thus the effect of this assumption is to bias the results of our simulation in the direction of greater consumption poverty alleviation, since we are adding earnings but not subtracting transfers that might be lost as a result.[8]

Once the employment and income simulation is complete, the hours of household production of individuals need to be estimated for all households that contain job recipients. The recipient pool contains all adults living in households that contain at least one job recipient. The donor pool contains all adults living in households in which all eligible adults are engaged in employment. The final step is imputing new consumption expenditures for the households that included job recipients. Once all these steps have taken place, we can recalculate LIMTCP using

---

[6] Eligible adults are defined as all individuals between the ages of 18 and 70 who are not disabled, retired, or in school.

[7] An exception will be noted in the discussion of the labor force simulation.

[8] The average total transfers for individuals in adjusted consumption poor households receiving transfers is 260 Turkish Lira per month, compared to the average adjusted poverty line for such individuals of 926 Turkish Lira per month.

the imputed values for time use and consumption expenditures. We now describe the method for each step in detail.

Labor Force Simulation

This simulation follows the method developed in prior research on time and income poverty, which built on research done at the Levy Institute to estimate the impact of the American Recovery and Reinvestment Act of 2009 on U.S. income inequality. The problem here is to assign hours and earnings to individuals receiving paid employment. The method for assigning hours and earnings is a hot-decking procedure (for a review of hot-decking, see Andridge and Little 2010). We use a nearest-neighbor method called affinity scoring to get a pool of records from which to match each record within matching cells determined by age, sex, and education. Before the hot-decking, we assign an industry and occupation to each job recipient. We also generate imputed wages and hours of work using a three-stage Heckit procedure. These four variables are used in the hot-decking assignment of hours and earnings. In addition to hours and earnings we assign industry, occupation and employment type (formal or informal).

Industry and Occupation Assignment

The first step in assigning jobs to recipients is to determine what are the likeliest industry and occupation for each of the recipients. This is done using a multinomial logit procedure. Both industry and occupation are regressed on age, sex, marital status, education and relationship to household head in the donor pool. The likelihood for each industry and occupation is then predicted in the recipient pool, using the results of the multinomial logit. Then each recipient is assigned the likeliest industry and occupation using those predicted likelihoods.

Imputed Hours and Earnings

The imputations for the earnings and usual weekly hours of paid work were performed using a three-stage Heckit procedure (Berndt 1991, p. 627). The model, described below, was run separately for each combination of six age categories and sex. The first stage is a probit estimation of labor force participation:

$$lf_i = \alpha_1 + \beta X + \varepsilon_i \tag{1}$$

The vector of explanatory variables, X, includes indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in

the household, number of children in the household, education, marital status, and spouse's age and education. The regression is run on the universe of all eligible adults. The Mills ratio is calculated using the results of the first stage regression:

$$\lambda = f(\frac{-\hat{lf}}{\hat{\sigma}_{\hat{lf}}}) \bigg/ \left(1 - F(\frac{-\hat{lf}}{\hat{\sigma}_{\hat{lf}}})\right) \tag{2}$$

Where f is the normal density function, F is the normal distribution function, $\hat{lf}$ is the estimated probability of labor force participation, and $\sigma_{\hat{lf}}$ is the standard deviation of $\hat{lf}$.

The second stage is an OLS estimate of the log of hourly wage:

$$\ln w_i = \alpha_2 + \gamma_2 Z + \theta_2 \lambda + \mu_i \tag{3}$$

The regression is run only on those that are actually employed for pay. The vector of explanatory variables, Z, in this stage includes the individual's education, age, marital status, industry and occupation, and finally, λ, the Mills Ratio calculated in the first stage. Inclusion of the Mills Ratio corrects for the selection bias induced by limiting the regression to those in paid employment. The imputed log of wage is predicted for donors and recipients from the results of the regression, with industry and occupation replaced for the latter by the assigned industries and occupations from each scenario.

The third stage is a regression of hours per week:

$$h_i = \alpha_3 + \gamma_3 Z + \omega \ln \hat{w}_i + \theta_3 \lambda + \eta_i \tag{4}$$

The regression is once again run only on those in paid employment. The vector of explanatory variables, Z, in this stage is the same as the previous stage, with the addition of indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, and the number of children in the household. Finally, the imputed wage predicted in the second stage and the Mills Ratio calculated in the first stage are included. Imputed hours per week are predicted for donors and recipients using the results of the regression, replacing the industry and occupation of the latter with their assigned values. The results of the last two stages give us the remaining variables with which we perform the hot-decking procedure to assign earnings, hours, industry, occupation and employment type.

8

Jobs Assignment

We can now assign earnings, usual hours of work, industry, occupation and employment type to those individuals in the recipient pool. The assignment method is statistical matching with hot-decking. The matches are performed within cells formed from combinations of age, sex and educational attainment. In some cases, in which there were no donors in a cell, cells were combined. The variables used to assess nearness of match are rural/urban status, family type, marital status, spouse's labor force status, assigned industry and occupation, indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, number of children and the two imputed variables, log of wage and hours worked. Rural/urban status was assigned a weight in the matching procedure large enough to assure that no rural-urban matches occurred. Industry and occupation are the next most heavily weighted variables. Next are imputed hours and wage, followed by family type type and then the variables relating to presence of children in the household. The selection of matches is done using affinity scoring.

Once the hot-decking is finished, we compare new earnings to previous earnings. In this employment simulation, there were a small number of individuals who actually reported earnings and who ended up with simulated earnings that were lower than their actually reported earnings. We removed these records from the pool of recipients and left their employment-related data unchanged. For the remaining recipients, we revised their household income by adding the total of the difference between the imputed amount of earnings and the actually reported earnings in the household (i.e., the sum of earnings differences of all recipients in the household) to the pre-simulation amount of household income.

Time Use Reallocation

We assume that as a result of the job assignment, the time use pattern of each eligible individual in the households that contain one or more job recipients (as adjusted) from the first stage will change. All adults in the recipient households are considered "eligible" for time-use reallocation. We use a second round of hot-decking to assign new weekly hours of household production to each of the "eligible" individuals, based on updated labor force participation variables for the recipients of jobs in the first stage. The method is the same as the first stage, with the exception of the matching variables used and their relative weighting in the procedure. In this stage, the variables used to assess nearness of match are rural/urban status, family type,

marital status, labor force status, spouse's labor force status, indicators for the presence of male and female children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, number of children in the household, number of adults in the household, household income, the income share of each individual,[9] and the two imputed variables from the first stage: earnings and usual weekly hours of employment. Household income and labor force status are updated to reflect the increased earnings and the new job assignments received in the previous stage. Again, rural/urban status is weighted to ensure no intra-regional matches occur. The number of children and number of adults in the household, household income, and income share are the next most heavily weighted variables. After that, the five employment-related variables assigned in the previous hot-decking step, marital status and spouse's labor force status, then the variables relating to children in the household. For each match, the weekly hours of household production are transferred. We now have the time use variables necessary to recalculate time and consumption poverty, but we still need to adjust household consumption expenditures to reflect the new, higher household incomes of recipient households.

Consumption Expenditures Adjustment

In order to estimate the change in consumption expenditures for recipient households we do a third hot-decking procedure. We first estimate household consumption expenditures using a tobit regression (with a zero lower bound) on the natural log of actual consumption expenditures reported in the HBA 2006. As explanatory variables we use the natural log of equivalence-scale adjusted household income, the number of children, the number of adults, the class of worker for head and spouse, the age, sex, and labor force status of household head. We run the regression separately for rural and urban households and predict log of consumption expenditures for all households.

We then do hot-decking, this time at the household level. The matching variables are rural/urban status, family type, marital status of the household head, assigned or actual employment type, industry, and occupation of the household head, the spouse of the household head's assigned or actual employment type, the imputed log of consumption expenditures, the log of equivalence-scale adjusted household income, the number of kids, and the number of

---

[9] Income share is included to reflect changes in bargaining power within the household and its impact on the distribution of household production work.

adults. Again, the rural/urban status variable was appropriately weighted to prevent rural-urban matches. The next most heavily weighted variable was the imputed consumption expenditures, followed by household income. The next most heavily weighted variables were the number of children and the number of adults, followed by marital status and family type.

For each match we transferred household monthly consumption expenditures. We compared the transferred expenditures to the actually-reported expenditures and in those cases in which the latter were higher we replaced the matched value with the actually reported consumption expenditures plus the increase in household income as a result of the increased earnings.[10] We move now to an assessment of the quality of the simulation results.

Quality Assessment

Assessing the quality of this type of simulation is difficult since we are producing a counterfactual distribution of earnings, time use and consumption expenditures. The assessment is therefore limited to comparing the latter qualities among sub-groups of donor and recipient records.

First we compare the recipient and donor pools for the first stage in the simulation. Figure 3 presents the breakdown of the recipients and donors by matching cell (based on sex, age and education). We can see that among women, the members of both donor and recipient pools tend to be on the younger side, while this trend is less pronounced among men. There are much fewer elderly women in either pool than men, and very few of the donors in the oldest category for either men or women. In all age and sex categories, the donors tend to be more highly-educated than the recipients. To a certain extent, the unbalance in the donor and recipient pools will tilt the results of the simulation, especially in the cases where there are significantly fewer donors than recipients.

We can compare the industry and occupation assigned in the employment simulation to the likeliest industry and occupation calculated in the first step of our procedure. This comparison is presented in Tables 7 and 8. As we can see, the assignment matched the likely industry in 92 percent of cases, while for occupation the match rate was over 95 percent. Assessing the earnings imputation is less straightforward, although we can compare the assigned earnings to actual earnings by matching cell for some indication. Figure 4 displays the ratio of

---

[10] We estimated the average propensity to consume for households below the adjusted consumption poverty line and found it to be above unity, on average.

mean and median assigned monthly earnings to actual monthly earnings for each combination of sex, age and education. The shaded area represents a band of plus or minus 20 percent from equivalence, a sweet spot. Many cells were empty of either donors or recipients and so have no ratios. The worst ratio is that for women aged 45 to 54 with a middle school education. Fortunately, this group represents only 14,000 of the 5.6 million recipients in the simulation. Generally, the more populated a cell with donors and recipients, the better the results of the simulation. Figure 5 displays the same ratios for usual weekly hours of work. The results here are clearly superior. It is intuitively obvious that it should be so since there is much lower variation in weekly hours of paid work than in earnings.

Turning to the estimation of weekly hours of household production, we again first compare the recipient and donor pools. Figure 6 shows the comparison by matching cell. We again see that many of the recipients and, to a lesser degree, the donors are younger (in the first two age categories) and less educated. This is again more pronounced for women in terms of education. Figure 7 presents a comparison based on sex, number of children in the household and number of adults. Here we see that more of the recipients are in households with two or three and more children, while most of the donors are from households with one or no children. This reflects somewhat the nature of the simulation and the social structure in Turkey, in that more of the eligible adults that are not working are to be found in households with more children, working, in fact, as household production workers. This will tend to have some impact on the assignment of weekly hours of household production.

Figures 8 and 9 contain ratios of recipients' mean and median assigned weekly hours of household production to donors actual mean and median hours, again by matching cells and by sex, number of children and number of adults, respectively. The results show that the distribution of assigned weekly hours by matching cell resembles the actual distribution of the donor pool. The cases that are the furthest from equivalence are among elderly men and women, and these cells were, again, lightly populated. For example, the worst case was for males aged 55 to 64 with college educations. This cell contained 1,300 recipients (compared to 11.2 million in the entire simulation). The comparison by sex and household composition is even better-looking.

The final assessment we do is of the imputation of consumption expenditures. Table 9 contains the actual and simulated mean and median consumption expenditures for rural and urban. As we would expect the mean and median for the simulation is higher than for the actual

reported consumption expenditures, since we are adding income to many of the adjusted consumption poor households.

To the best of our ability to judge, the simulation looks like a reasonable approximation of the impact on individual adjusted consumption-poor households of all eligible adults acquiring paid employment. The results of the simulation will tend to give an optimistic view of the impact of such employment transitions, since we cannot account for the loss of transfers. Nevertheless, the results should shed an interesting light on the effectiveness of employment generation as a consumption-poverty alleviation policy in Turkey.

REFERENCES

Andridge, Rebecca R., and Roderick J. A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78 (1): 40-64. doi:10.1111/j.1751-5823.2010.00103.x.

Berndt, Ernst R. 1996. *The Practice of Econometrics: Classic and Contemporary*. Har/Dis. Addison Wesley, March 14.

Kum, Hyunsub, and Thomas Neal 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, no. 3 (January 1): 177-196. doi:10.3233/JEM-2010-0332.

Masterson, Thomas. 2010. "Quality of Match for Statistical Matches Used in the 1992 and 2007 LIMEW Estimates for the United States." Levy Economics Institute Working Paper 618 (September). http://www.levyinstitute.org/pubs/wp_618.pdf.

Masterson, Thomas. 2012. "Simulations of Full-Time Employment and Household Work in the Levy Institute Measure of Time and Income Poverty (LIMTIP) for Argentina, Chile, and Mexico." Levy Economics Institute Working Paper 727 (July). http://www.levyinstitute.org/pubs/wp_727.pdf

Zacharias, Ajit, Thomas Masterson, and Kijong Kim. 2009. "Distributional Impact of the American Recovery and Reinvestment Act: A Microsimulation Approach." Levy Economics Institute Working Paper 568 (June). http://www.levyinstitute.org/pubs/wp_568.pdf

TABLES

Table 1 Alignment of Strata Variables, Turkey

	HBA2006	ZKA2006	Difference
Number of children in household			
0	32.7%	35.9%	-3.2%
1	23.7%	23.9%	-0.3%
2	25.3%	22.8%	2.4%
3+	18.4%	17.4%	1.0%
Number of adults in household			
1	5.2%	8.3%	-3.1%
2	54.8%	58.3%	-3.5%
3+	40.0%	33.4%	6.6%
Non-employed adult in household (y/n)?			
No	16.2%	15.9%	0.3%
Yes	83.8%	84.1%	-0.3%
Within 75 and 150% of poverty line (y/n)?			
No	62.7%	60.7%	2.0%
Yes	37.3%	39.4%	-2.0%
Household income categories (ZKA)			
300 YTL or less	4.6%	11.0%	-6.3%
301-450 YTL	7.5%	16.3%	-8.8%
451-600 YTL	13.2%	18.6%	-5.4%
601-750 YTL	10.3%	11.1%	-0.8%
751-1000 YTL	17.5%	17.9%	-0.3%
1001-1250 YTL	13.1%	7.0%	6.1%
1251-1750 YTL	16.0%	8.3%	7.7%
1751-2500 YTL	10.7%	6.0%	4.7%
2501-4000 YTL	5.1%	2.8%	2.3%
4001 YTL or more	2.0%	1.0%	1.0%
Employed (y/n)?			
No	55.0%	53.6%	1.4%
Yes	45.0%	46.4%	-1.4%
Sex			
Male	47.6%	49.2%	-1.5%
Female	52.4%	50.9%	1.5%

Table 2 Distribution of Matched Records by Matching Round, Turkey

Round	Matched Individuals	Percent of Total	Cumulative Percentage
1	33,690,418	65.2%	65.2%
2	2,243,594	4.3%	69.5%
3	328,500	0.6%	70.2%
4	157,563	0.3%	70.5%
5	8,062,717	15.6%	86.1%
6	35,379	0.1%	86.2%
7	42,674	0.1%	86.2%
8	1,478,360	2.9%	89.1%
9	8,462	0.0%	89.1%
10	158,738	0.3%	89.4%
11	90,872	0.2%	89.6%
12	107,388	0.2%	89.8%
13	121,944	0.2%	90.0%
14	41,812	0.1%	90.1%
15	286,887	0.6%	90.7%
16	158,004	0.3%	91.0%
17	183,466	0.4%	91.3%
18	1,162,931	2.3%	93.6%
19	17,685	0.0%	93.6%
20	13,955	0.0%	93.6%
21	194,706	0.4%	94.0%
22	52,653	0.1%	94.1%
23	30,121	0.1%	94.2%
24	19,441	0.0%	94.2%
25	188,819	0.4%	94.6%
26	55,853	0.1%	94.7%
27	214,124	0.4%	95.1%
28	42,682	0.1%	95.2%
29	66,140	0.1%	95.3%
30	117,826	0.2%	95.5%
31	105,853	0.2%	95.8%
32	26,632	0.1%	95.8%
33	119,141	0.2%	96.0%
34	88,038	0.2%	96.2%
35	77,711	0.2%	96.4%
36	30,456	0.1%	96.4%
37	16,080	0.0%	96.4%
38	3,749	0.0%	96.5%
39	8,951	0.0%	96.5%
40	338,532	0.7%	97.1%
41	54,740	0.1%	97.2%
42	144,229	0.3%	97.5%
43	20,314	0.0%	97.5%
44	22,487	0.0%	97.6%
45	21,455	0.0%	97.6%
46	652,585	1.3%	98.9%
47	142,935	0.3%	99.2%
48	42,560	0.1%	99.3%
49	55,552	0.1%	99.4%
50	37,905	0.1%	99.4%
51	7,870	0.0%	99.5%
52	283,195	0.5%	100.0%
Total	51,674,685		

Table 3 Distribution of Weekly Hours of Household Production in ZKA 2006 and Matched File

	p90/p10	p90/p50	p50/p10	p75/p25	p75/p50	p50/p25	Gini
ZKA 2006	.	3.96	.	16.50	2.75	6.00	0.5521
Matched File	.	3.92	.	16.75	2.79	6.00	0.5519

Table 4 Comparison of Mean and Median Time Use Variables in Matched File to ZKA 2006

Average	**Core**	**Procurement**	**Care**	**Household Production**
MATCH	16.76	1.64	3.54	21.95
ZKA2006	16.68	1.65	3.58	21.92
Ratio	100.48%	99.39%	98.88%	100.14%
Median	**Core**	**Procurement**	**Care**	**Household Production**
MATCH	7.58	0.00	0.00	14.00
ZKA2006	7.58	0.00	0.00	14.00
Ratio	100.00%			100.00%

Table 5 Mean and Median Weekly Hours of Household Production by Strata Variable, ZKA 2006 and Matched File

Mean Weekly Hours of Household Production

	MATCH	ZKA2006	Ratio		MATCH	ZKA2006
Core	16.76	16.68	100.5%			
Procurement	1.64	1.65	99.4%			
Care	3.54	3.58	98.9%			
Household Production	21.95	21.92	100.1%			
Number of Children					**MATCH**	**ZKA2006**
0 children	19.58	19.06	102.7%			
1 child	21.17	21.33	99.2%	**1/0**	1.08	1.12
2 children	23.43	24.10	97.2%	**2/0**	1.20	1.26
3 or more children	24.47	24.91	98.2%	**3+/0**	1.25	1.31
Number of Adults						
1 adult	24.85	24.02	103.5%			
2 adults	25.10	24.74	101.5%	**2/1**	1.01	1.03
3 or more adults	19.50	19.33	100.9%	**3+/1**	0.78	0.80
Non-employed adult in household (y/n)						
No	18.09	18.87	95.9%	**yes/no**	1.25	1.19
Yes	22.60	22.39	100.9%			
Within Poverty Band (y/n)						
No	21.29	21.31	99.9%	**yes/no**	1.08	1.07
Yes	23.00	22.82	100.8%			
Household Income Category						
450 YTL or less	25.32	24.58	103.0%			
451-750 YTL	23.22	22.74	102.1%	**2nd/1st**	0.92	0.93
751-1250 YTL	22.07	21.00	105.1%	**3rd/1st**	0.87	0.85
1251-2500 YTL	20.73	19.17	108.1%	**4th/1st**	0.82	0.78
2501 YTL or more	18.59	17.10	108.7%	**Top/1st**	0.73	0.70
Employed (y/n)						
No	30.95	31.30	98.9%	**yes/no**	0.35	0.35
Yes	10.95	11.08	98.8%			
Sex						
Male	6.49	6.11	106.2%	**Fem/Male**	5.55	6.09
Female	35.99	37.20	96.7%			
Rural/Urban						
Urban	21.90	21.17	103.4%	**Rur/Urb**	1.01	1.09
Rural	22.03	23.18	95.0%			

Median Weekly Hours of Household Production

	MATCH	ZKA2006	Ratio		MATCH	ZKA2006
Core	7.58	7.58	100.0%			
Procurement	0.00	0.00				
Care	0.00	0.00				
Household Production	14.00	14.00	100.0%			
Number of Children					MATCH	ZKA2006
0 children	13.42	13.42	100.0%			
1 child	12.25	11.67	105.0%	1/0	0.63	0.61
2 children	14.58	14.58	100.0%	2/0	0.74	0.76
3 or more children	15.75	15.75	100.0%	3+/0	0.80	0.83
Number of Adults						
1 adult	22.75	21.58	105.4%			
2 adults	18.08	16.92	106.9%	2/1	0.73	0.70
3 or more adults	10.50	10.50	100.0%	3+/1	0.42	0.44
Non-employed adult in household (y/n)						
No	11.67	12.83	91.0%	yes/no	1.25	1.09
Yes	14.58	14.00	104.1%			
Within Poverty Band (y/n)						
No	13.42	13.42	100.0%	yes/no	1.09	1.04
Yes	14.58	14.00	104.1%			
Household Income Category						
450 YTL or less	20.42	18.08	112.9%			
451-750 YTL	15.75	15.17	103.8%	2nd/1st	0.62	0.62
751-1250 YTL	13.42	12.25	109.6%	3rd/1st	0.53	0.50
1251-2500 YTL	12.25	10.50	116.7%	4th/1st	0.48	0.43
2501 YTL or more	10.50	9.33	112.5%	Top/1st	0.41	0.38
Employed (y/n)						
No	31.50	31.50	100.0%	yes/no	0.15	0.15
Yes	4.67	4.67	100.0%			
Sex						
Male	2.92	2.92	100.0%	Fem/Male	12.59	12.78
Female	36.75	37.33	98.4%			
Rural/Urban						
Urban	14.00	12.83	109.1%	Rur/Urb	0.96	1.18
Rural	13.42	15.17	88.5%			

Table 6 Ratio of Matched to ZKA 2006 Average Hours of Household Production for the Reference Groups

Number of Children	Number of Adults		
	1	2	3+
0	101.8%	111.3%	99.0%
1	104.1%	114.5%	109.3%
2	111.3%	107.5%	106.7%
3+	115.6%	112.9%	100.8%

Table 7 Likely and Assigned Industries for Labor Market Simulation Recipients

Assigned Industry	Likely Industry						
	Agriculture, hunting, forestry, fishing	Manufacturing	Construction and public works	Wholesale and retail trade, hotels and restaurants	Financial intermediation, real estate, renting and business activities	Community, social and personal service activities	Total
Agriculture, hunting, forestry, fishing	**331,353**	23,843	11,067	-	-	179,450	545,713
Manufacturing	-	**1,924,960**	1,069	-	-	56,254	1,982,283
Construction and public works	-	3,005	**61,320**	-	-	20,771	85,096
Wholesale and retail trade, hotels and restaurants	35,554	20,733	2,732	**178,774**	3,679	14,519	255,991
Financial intermediation, real estate, renting and business activities	-	7,920	-	-	**-**	-	7,920
Community, social and personal service activities	14,388	62,908	5,323	-	-	**2,722,178**	2,804,797
Total	381,295	2,043,369	81,511	178,774	3,679	2,993,172	5,681,800
Percent Match	**86.9%**	**94.2%**	**75.2%**	**100.0%**	**0.0%**	**90.9%**	**91.8%**

Table 8 **Likely and Assigned Occupations for Labor Market Simulation Recipients**

Assigned Occupation	Likely Occupation								
	Legislators, senior officials and managers	Professionals	Associate professionals	Office and customer support workers	Service and sales workers	Craft and related trades workers	Plant and machine operators and assemblers	Laborers	Total
Legislators, senior officials and managers	**-**	-	-	-	1,069	-	2,395	-	3,464
Professionals	-	**44,912**	-	1,364	-	-	-	8,662	54,938
Associate professionals	-	-	**4,988**	-	2,078	-	-	-	7,066
Office and customer support workers	-	-	-	**122,182**	3,637	-	-	-	125,819
Service and sales workers	2,179	-	-	6,090	**410,416**	13,836	-	20,718	453,239
Craft and related trades workers	-	-	-	2,691	56,898	**847,729**	-	75,942	983,260
Plant and machine operators and assemblers	-	-	-	-	-	643	**46,892**	29,816	77,351
Laborers	-	-	-	-	10,167	24,417	-	**3,942,079**	3,976,663
Total	2,179	44,912	4,988	132,327	484,265	886,625	49,287	4,077,217	5,681,800
Percent Match	**0.0%**	**100.0%**	**100.0%**	**92.3%**	**84.8%**	**95.6%**	**95.1%**	**96.7%**	**95.4%**

21

Table 9 Actual and Simulated Mean and Median Household Consumption Expenditures by Rural/Urban Status (Turkish Lira)

		Actual	**Simulated**
Urban	**Mean**	1,208	1,325
	Median	1,053	1,174
Rural	**Mean**	885	1,025
	Median	756	901

Figure 1 Ratio of Mean Household Production by Category (Match/ZKA 2006)

	Number of kids	Number of adults	In Poverty Band	Non-emp. Adult in HH	HH Income	Employed	Sex	Overall
■ cat1	102.7%				103.0%			
■ cat2	99.2%	103.5%	99.9%	95.9%	102.1%	98.9%	106.2%	100.1%
■ cat3	97.2%	101.5%	100.8%	100.9%	105.1%	98.8%	96.7%	100.0%
■ cat4	98.2%	100.9%			108.1%			
■ cat5					108.7%			
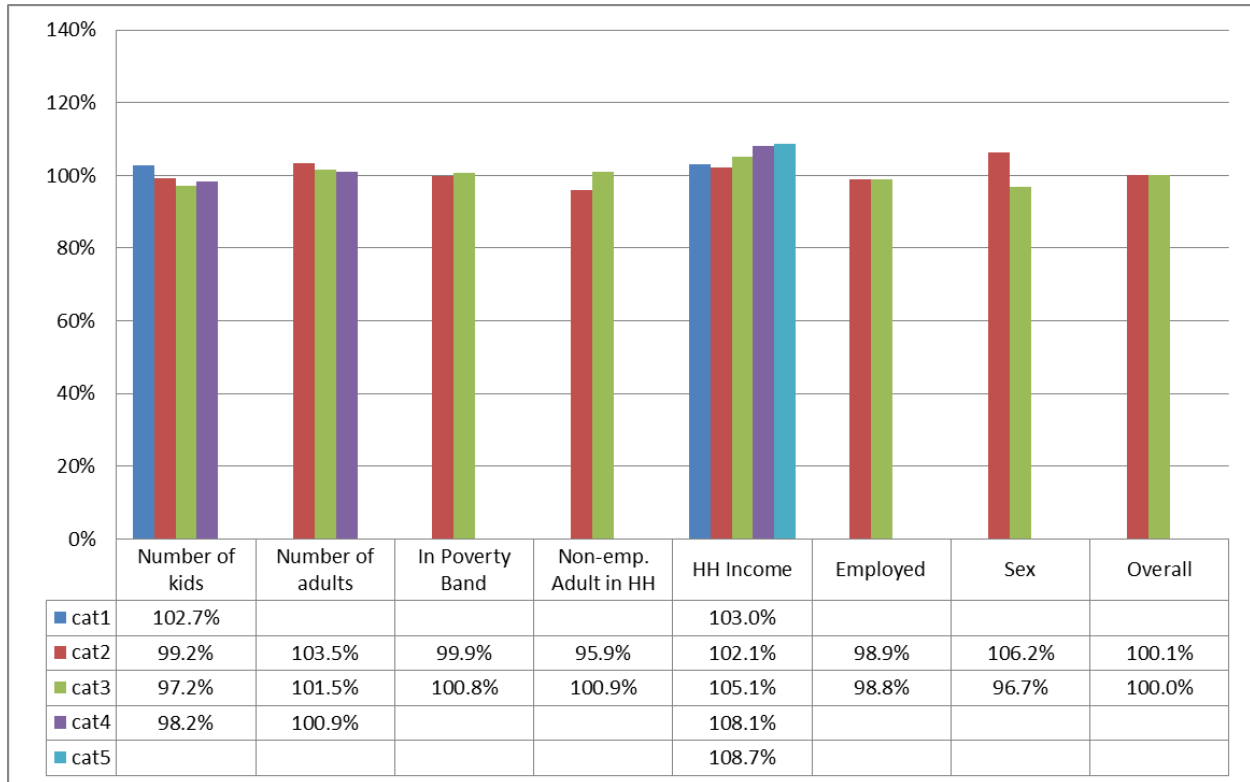
Figure 2 Household Production by Reference Groups, ZKA 2006 and Matched File

Ad=Adult, Ch=Child

Weekly Hours of HH Production

Graphs by Dataset
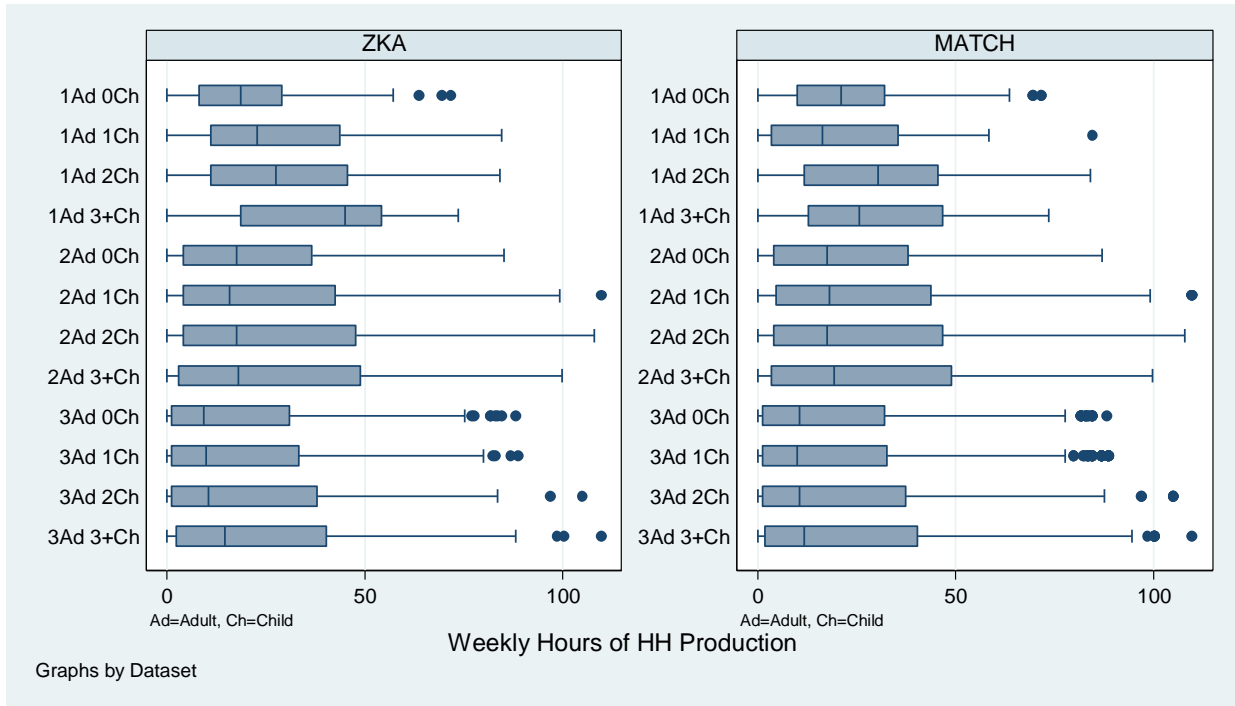
Figure 3 Donor and Recipient Pools for Labor Force Simulation by Sex, Age and Education
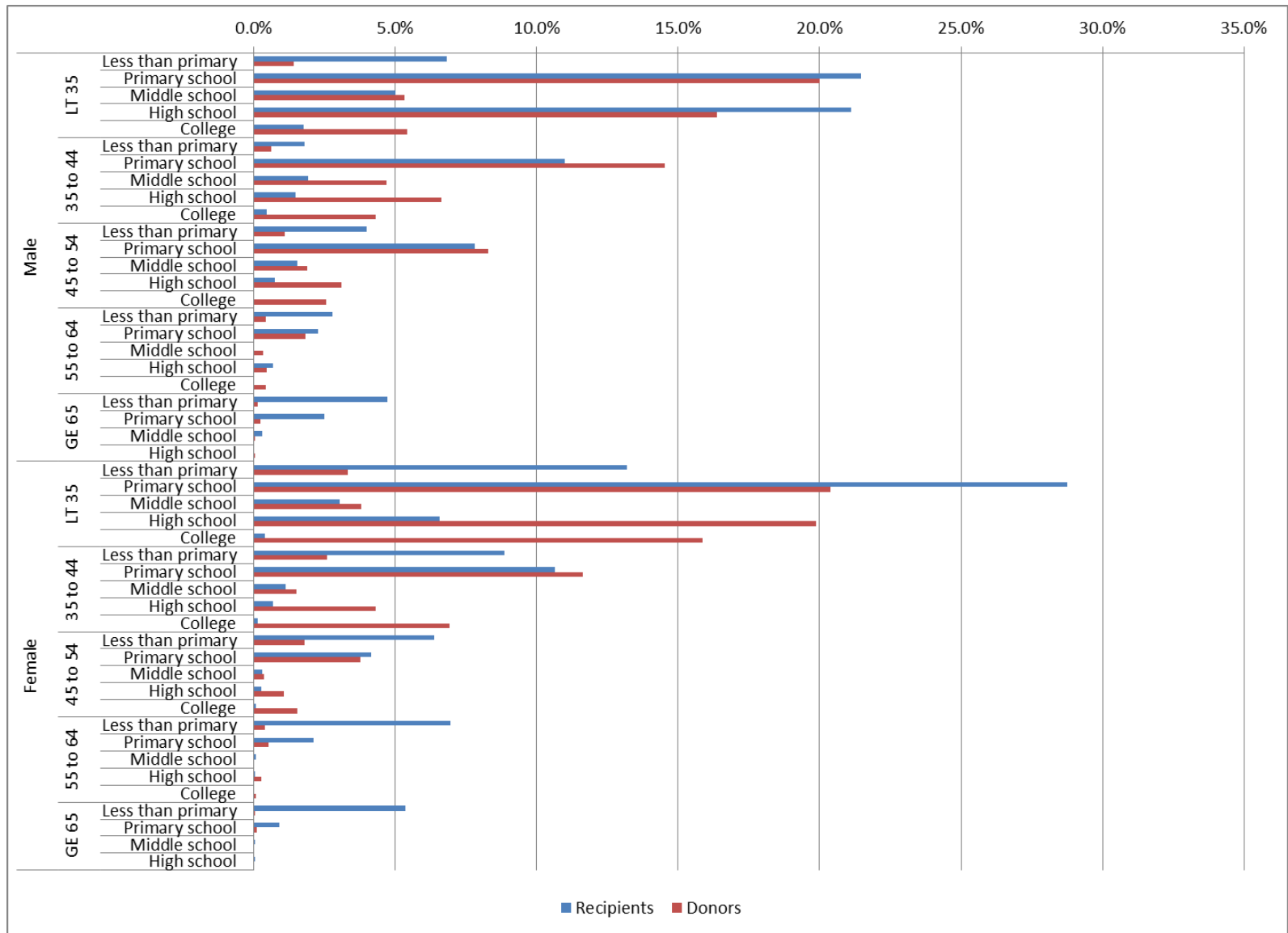
Figure 4 Mean and Median Earned Income for Donors and Recipients for Labor Market Simulation by Sex, Age and Education
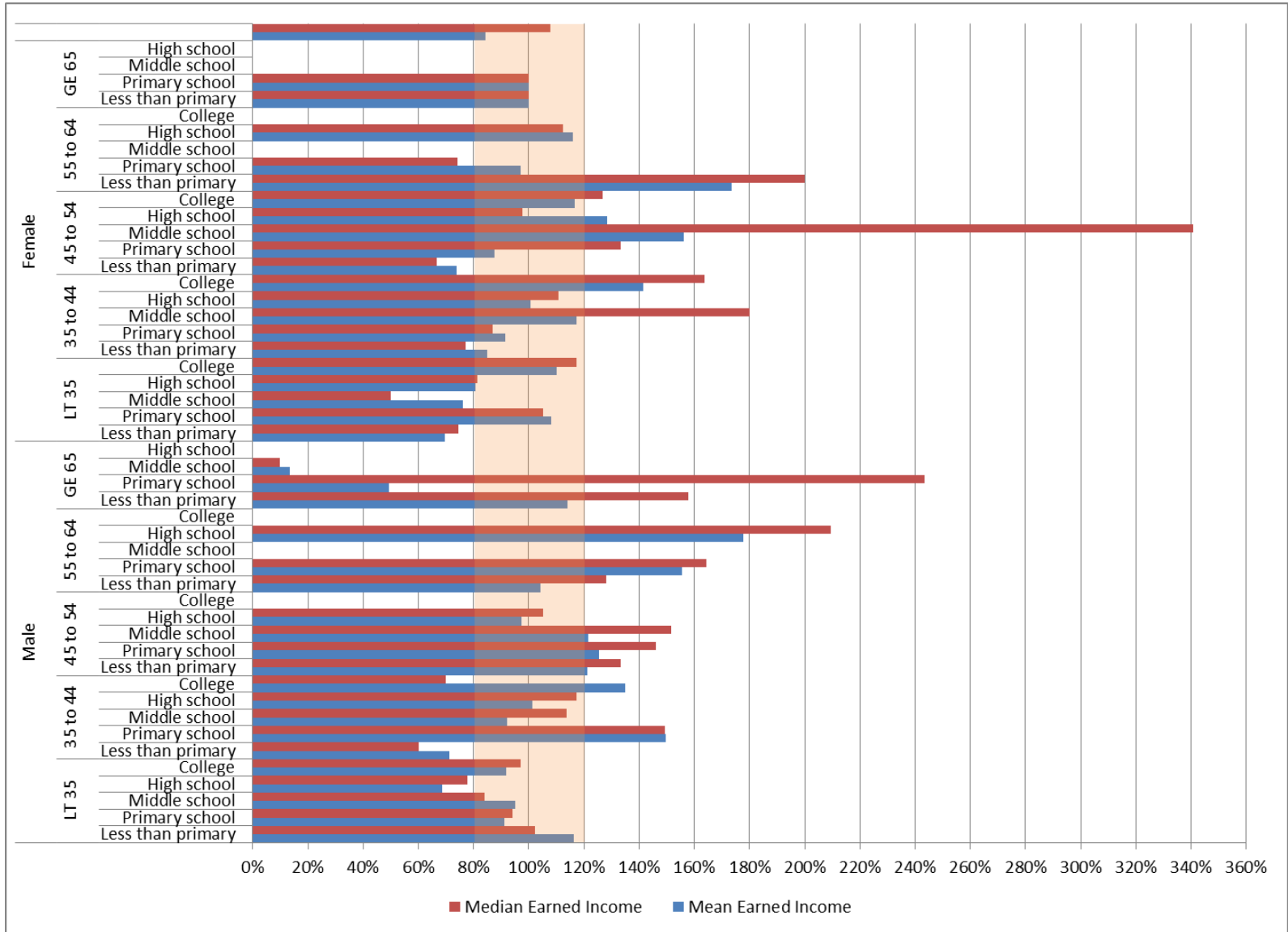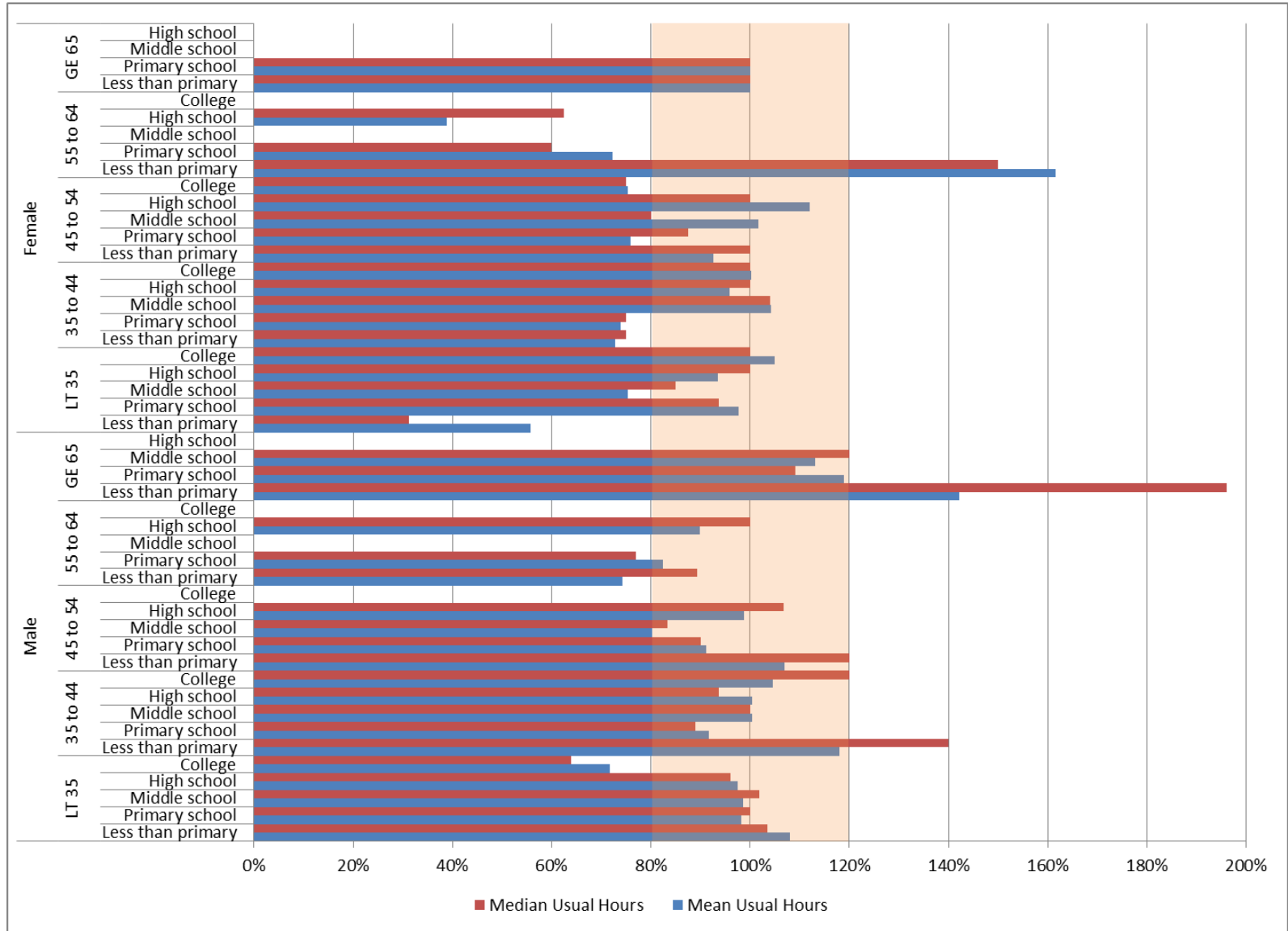
Figure 5 Mean and Median Usual Hours of Work for Donors and Recipients for Labor Market Simulation by Sex, Age and Education

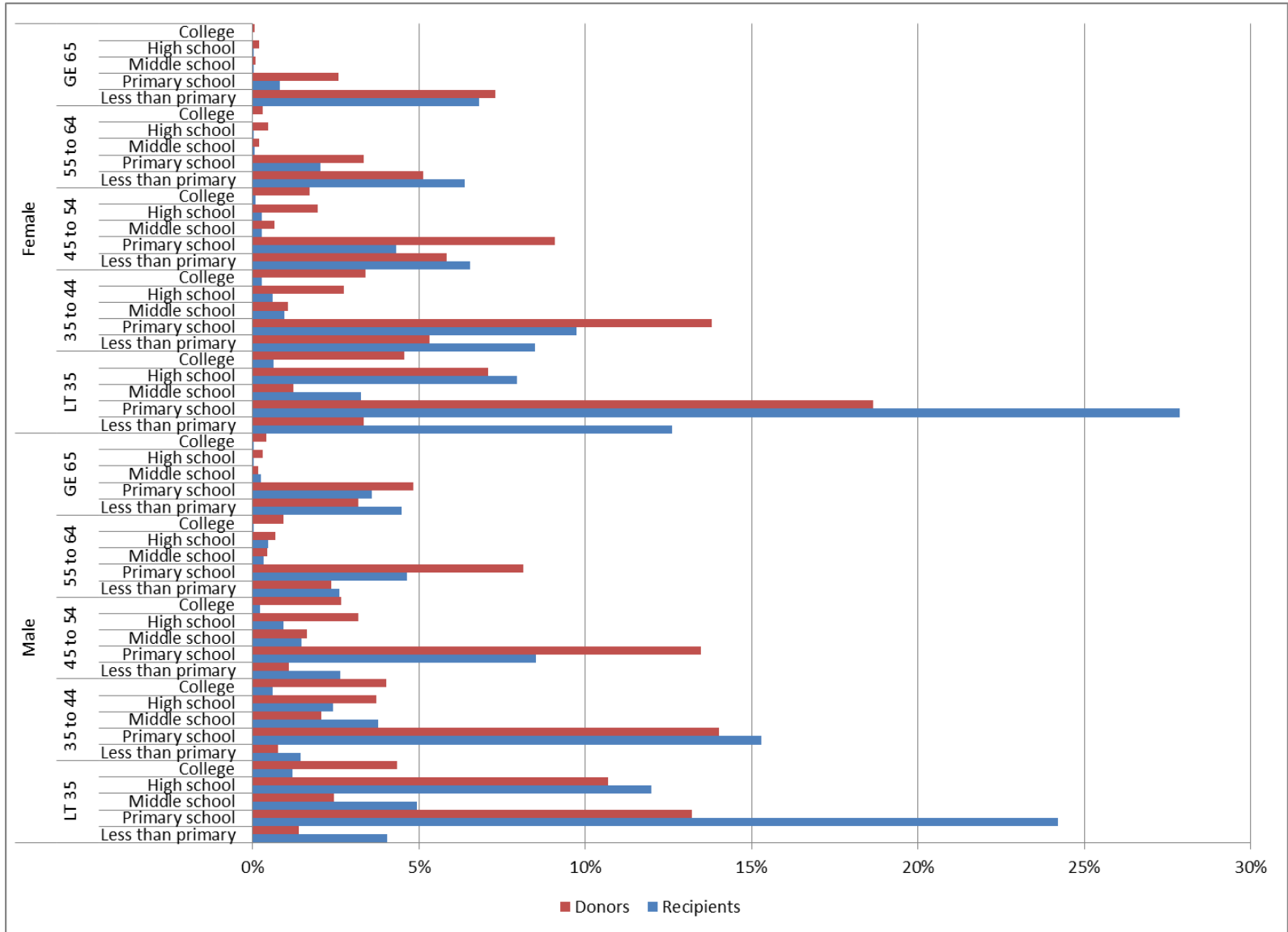*Figure 6* Donor and Recipient Pools for Time Use Simulation by Sex, Age and Education

Figure 7 Donor and Recipient Pools for Time Use Simulation by Sex, Number of Adults and Number of Children
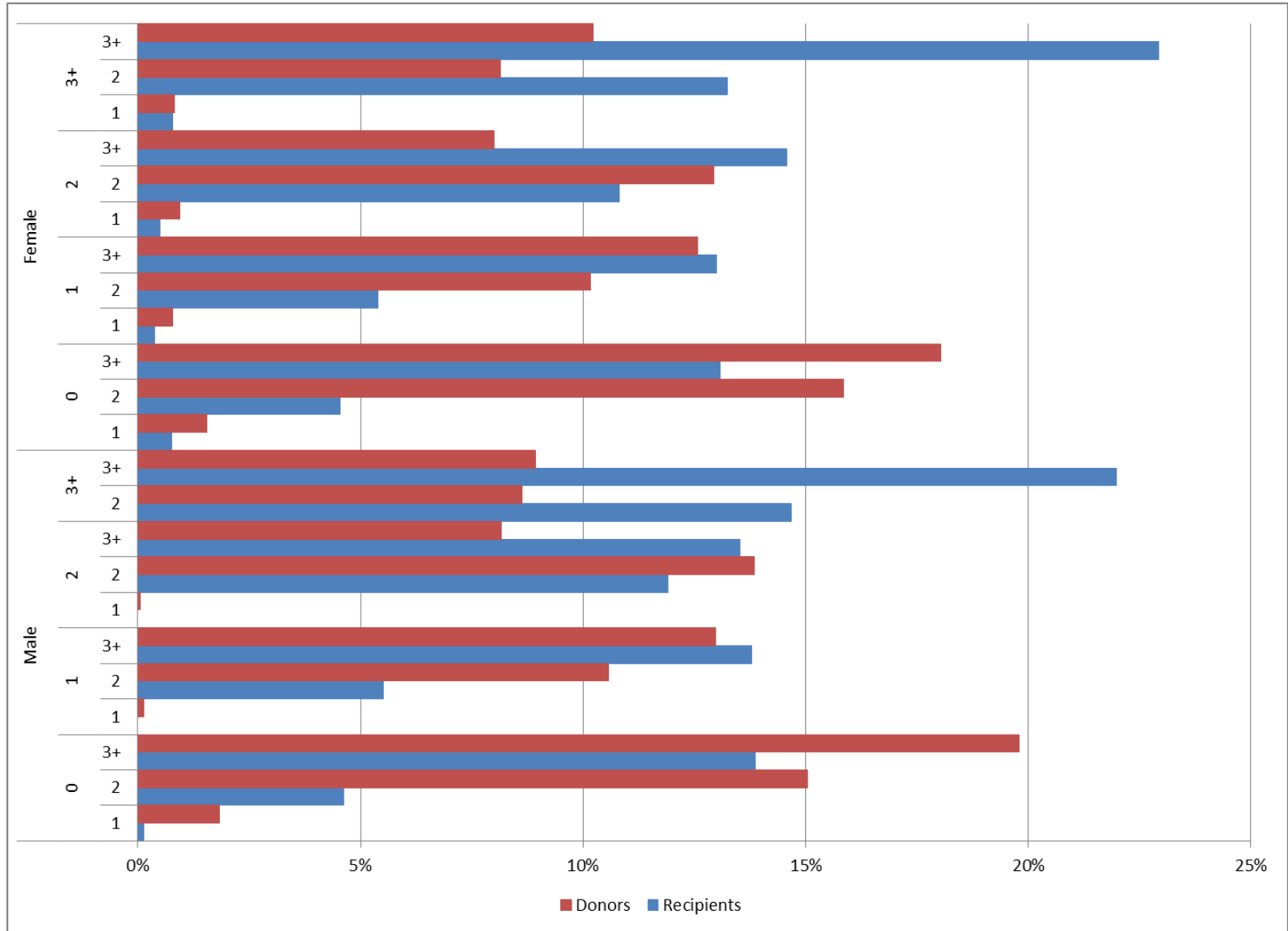
Figure 8 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Time Use Simulation by Sex, Age and Education
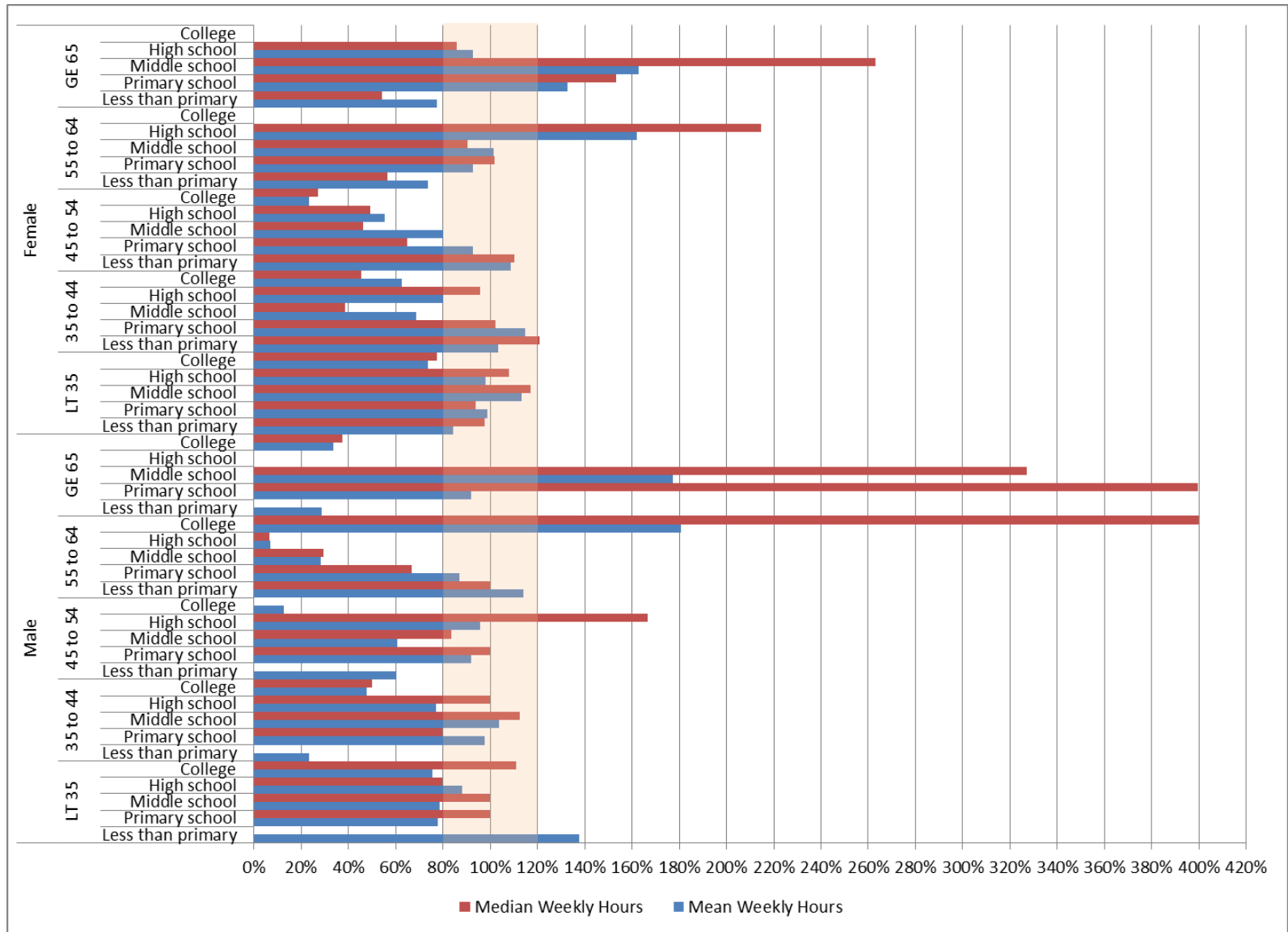
Figure 9 Mean and Median Weekly Hours of Household Production for Donors and Recipients for Time Use Simulation by Sex, Number of Adults and Number of Children