

Masterson, Thomas

**Working Paper**

Quality of statistical match and employment simulations used in the estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for South Korea, 2009

Working Paper, No. 793

**Provided in Cooperation with:**

Levy Economics Institute of Bard College

*Suggested Citation:* Masterson, Thomas (2014) : Quality of statistical match and employment simulations used in the estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for South Korea, 2009, Working Paper, No. 793, Levy Economics Institute of Bard College, Annandale-on-Hudson, NY

This Version is available at:

<https://hdl.handle.net/10419/109986>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



## Working Paper No. 793

---

### **Quality of Statistical Match and Employment Simulations Used in the Estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for South Korea, 2009**

by

**Thomas Masterson**

Levy Economics Institute of Bard College

**March 2014**

---

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>

Copyright © Levy Economics Institute 2014 All rights reserved

ISSN 1547-366X

## **ABSTRACT**

The quality of match of the statistical match used in the LIMTIP estimates for South Korea in 2009 is described. The match combines the 2009 Korean Time Use Survey (KTUS 2009) with the 2009 Korean Welfare Panel Study (KWPS 2009). The alignment of the two datasets is examined, after which various aspects of the match quality are described. The match is of high quality, given the nature of the source datasets. The method used to simulate employment response to availability of jobs in the situation in which child-care subsidies are available is described. Comparisons of the donor and recipient groups for each of three stages of hot-deck statistical matching are presented. The resulting distribution of jobs, earnings, usual hours of paid employment, household production hours, and use of child-care services are compared to the distribution in the donor pools. The results do not appear to be anomalous, which is the best that can be said of the results of such a procedure.

**Keywords:** Statistical Matching; Time Use; Household Production; Poverty; LIMTIP; South Korea

**JEL Classifications:** C14, C40, D31, J22

## INTRODUCTION

This paper describes the construction of synthetic datasets created for use in estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for South Korea in 2009.<sup>1</sup> This work was carried out for a project contracted by the Korean Employment Information Service (KEIS).<sup>2</sup> Construction of LIMTIP 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 South Korea. Thus, in order to produce LIMTIP estimates, a synthetic data file is created by statistically matching two source data sets.<sup>3</sup> We use the Korean Welfare Panel Study (KWPS 2009), the unification of “the near-poor and the poor panel, self-support panel” by the Korean Institute for Health and Social Affairs and the “Korean welfare panel” by the Social Welfare Research Center of Seoul National University as the base data set, since it contains good information on demographics, income, transfers and taxes for a representative sample of households in South Korea. Time use data come from the Korean Time Use Survey (KTUS 2009), which is also nationally representative.

In order to assess the possible impact of income-poverty reduction strategies founded upon expanding employment on time and income poverty, it is necessary to impute the impact of those strategies on the income, time allocation, and childcare utilization of households. We draw on and extend our work simulating the results of the American Recovery and Reinvestment Act (Zacharias et al., 2009) and previous LIMTIP employment simulations (Masterson 2012, 2013). In this case, we assume that some unspecified way is found to employ those adults in households underneath our adjusted income poverty line who are not employed. We then assess the impact this change has on time and income poverty.

This paper is organized as follows. The source datasets are described and their demographic characteristics are compared. Then the quality of the statistical match is reviewed including diagnostics about the match itself. Next we describe the methodology involved in the imputation of occupation, industry, and employment type, hours of employment and earnings,

---

<sup>1</sup> For a description of the theory and methodology behind producing estimates of the Levy Institute Measure of Time and Income Poverty (LIMTIP), see Zacharias (2011).

<sup>2</sup> The project, titled “Employment and Social Policies for Time and Income Poor: application of LIMTIP in South Korea,” is managed by Tae-Hee Kwon of KEIS with co-investigators Ajit Zacharias, Rania Antonopoulos, Thomas Masterson and Kijong Kim of the Levy Economics Institute of Bard College.

<sup>3</sup> See Kum and Masterson (2010) for details of the statistical matching procedure that we use.

household income, household production hours, and childcare hours. Finally we assess the results of the simulation.

## **I. STATISTICAL MATCHING**

### **Data and Alignment**

The source data sets for the time use match for the LIMTIP estimates for South Korea are the 2009 KWPS and the 2009 KTUS. We use individual records from the 2009 KWPS file, excluding those living in group quarters or in the armed forces. Since the KTUS covers individuals aged 10 years and older, we discard younger individuals from the KWPS file. This leaves 14,502 records, which represents 43,219,236 individuals when weighted. In the KTUS, we have 20,263 individual records, representing 43,297,959 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 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% and 150% 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.

### **Estimation of the Poverty Band in the Time Use Micro Data**

In order to do the time use match required for the estimation of the LIMTIP for Korea, we need to be able to identify the individuals from households within the poverty band. This variable can

be directly calculated in the KWPS, but because the KTUS data has limited income variables (only categorical personal income), we have to impute a household's presence in the poverty band in the time use data. We do this by using the predicted probability of being within the poverty band by means of a probit estimation.

We begin by constructing a household income measure for households in the time use data. For each individual, we create a personal income variable using the midpoint of the categories of the existing personal income variable, and replacing the top category (over ₩5,000,000) with ₩6,000,000. The household income is then created by summing these across all members of the household. This results in a household income distribution in the time use data that has a substantially lower mean than that in the welfare data (₩2.6 million versus ₩3.5 million). We normalize the household income data in the welfare and time use data separately, in order to produce similar distributions for the probit estimation and prediction.

We then proceed to run probit estimations on each of the reference group categories for the required household production (12 combinations of number of adults —one to three or more— and number of children — zero to three or more— in the household) in the KWPS. The dependent variable is an indicator of presence in the poverty band, and the independent variables are standardized household income, number of persons in the household, a set of dummies for seven regions of the country, the sex of the household head, the age and square of age of the household head, dummies for family type, dummies for tenure status, dummies for the type of housing unit, the number of earners in the household, and the level of education of the household head. The results of the estimation are used to predict the presence of the household in the poverty band for all household records in both the time use and the welfare data. We estimate the latter in order to assess the quality of the procedure. The results for the procedure are presented in Table 1. As we can see, the rate of misprediction is quite low, at 8.5%. In addition, the highest income of those households in the welfare data that were miscategorized as being within the poverty band was ₩3.5 million, which is not too far above the maximum poverty line for welfare data of ₩2.2 million. This gives us confidence in our estimates, and the matching can proceed.

### **Alignment of the Time Use and Welfare Panel Surveys**

Table 2 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 would expect them to be

well aligned. Unfortunately, that seems not to be the case. We see that there are 3% fewer individuals in households without children in the KWPS than in the KTUS, while individuals in two-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 KTUS and those in three-or-more-adult households are more common in the KWPS, a difference of five percentage points. The ratio of individuals in households with at least one non-employed adult differs by 7% between the two surveys, while the ratio of individuals in households within the poverty band is 4% higher in the KTUS. The distribution by household income is noticeably more skewed to the lower end of the distribution in the KTUS compared to the KWPS. The portion of households in the lowest income category is six percentage points higher in the KTUS while the share in the highest income category is eight percentage points lower. This is due to the poor quality of the household income question and data in the time use survey. The nonemployed are under-represented in the KTUS relative to the KWPS (3%). The distribution of individuals by sex, at least, is close in the two surveys, with less than one half a percent separating the share of women in the KTUS and the KWPS. So, we have a relatively bad alignment between the two surveys compared to other statistical matches we have done.

### *Match QC*

Turning to the results of the match, we first look to the distribution of matched records by matching round in Table 3. The bulk of the matches, 69.5%, 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, and their relative mis-alignment.<sup>4</sup> The rest of the records are matched over an additional 16 rounds, with 1.7% receiving no match at all (Round 18). Table 4 provides a comparison of the distribution of weekly hours of household production in the KTUS and the matched file. The tenth percentile is zero, so two of the percentile ratios are undefined. The remaining percentile ratios are all relatively close, with the ratio of the 75<sup>th</sup> percentile to the median being exactly equivalent. The Gini coefficient is quite close, 0.625 in the matched file, compared to 0.627 in the KTUS. Table 5 breaks down the mean and median of the three categories of household production and the total in the matched file and

---

<sup>4</sup> 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 170.

the KTUS.<sup>5</sup> 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, being 2.8% but amounting to less than two minutes lower in the matched file than in the KTUS. The total is off by less than one percent, amounting to about eight and one half minutes. 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 KTUS for the seven strata variables. For almost all the categories, the average weekly hours in the matched file are within 5% of the KTUS. One exception is for males, who have 9% higher weekly hours in the matched file, although this amounts to a difference of only one half hour. The other is for individuals in households with at least one non-employed adult, who have about 7% lower weekly hours in the matched file. This amounted to a bit over one fewer hour per week. Overall, the ratios of the mean weekly household production hours in the matched file to those in the KTUS by the strata variables were quite close to one.

Table 6 has the actual numbers, and we can see that the large percentage differences represent relatively small differences in hours per week. For example, the largest percentage gap among income categories, that in-the-middle income category, we see that the actual amount of time difference is about 45 minutes per week. Notice also that the ratios by category are well reproduced in the matched file. The largest deviation is by presence of non-employed adults in the household, as we would expect given the differences in the averages individuals living in those households. 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 7.<sup>6</sup> We can see very little difference between the matched file and the KTUS in the distribution of weekly hours for individuals in Figure 2. Table 7 shows the ratio of household total hours of weekly production for households in the reference group in the matched file to the KTUS. Although the average values of weekly household production hours in the matched file as much as 38% lower than in the KTUS for some categories, those categories have relatively few households in them. Meanwhile, the ratios for the two adult households (the most numerous

---

<sup>5</sup> The three categories are care (child care, elder care, etc.), procurement (shopping, etc.), and core (cooking, cleaning, laundry, etc.).

<sup>6</sup> For the sake of clarity of the plot, only the number of children and number of adults is used.



groups) are all within ten percent. 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, but given the overall misalignment of the two source datasets, the results are quite good. 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.

## **II. LABOR MARKET SIMULATION**

### **Data and Methods**

The purpose of the simulation is to assess the first order impacts of policies aimed at alleviating income poverty in Korea via jobs policies, for example an employer of last resort (ELR) policy. In the case of Korea, substantial subsidies for childcare are used to promote women's labor force participation. These subsidies need to be taken into account in the estimation of time and income poverty. 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 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 levels of household total household production, as well as childcare hours, both privately paid for and subsidized, for the households with job recipients. Using these three steps, we can estimate the impact of a given policy on time and income 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 LIMTIP Korea project: that all eligible adults<sup>7</sup> in households below the adjusted income poverty line that are not working receive paid (either regular or irregular) employment.<sup>8</sup> The donor pool contains all adults 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

---

<sup>7</sup> Eligible adults are defined as all individuals between the ages of 18 and 54 who are not disabled, retired, or in school.

<sup>8</sup> An exception will be noted in the discussion of the labor force simulation.

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, so 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. Thus the effect of this assumption is to bias the results of our simulation in the direction of greater income poverty alleviation, since we are adding earnings but not subtracting transfers that might be lost as a result.<sup>9</sup>

Once the employment and income simulation is complete, the hours of household production of individuals needs to be estimated in 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 total household production hours in combination with child care hours contracted, both privately paid for and publically subsidized, for the households that included job recipients. The unit of analysis in this final step is the household itself. The donor pool contains all households that have all eligible adults working for pay. The recipient pool comprises all households with a recipient of a job in the first step. When we reassign total household production hours, we divide these hours up and assign them to individuals within the household using the shares of household production calculated with the results of the second step. Once all these steps have taken place, we can recalculate LIMTIP using the imputed values for time use, income, and child care services contracted. We now describe the method for each step in detail.

### **III. LABOR FORCE SIMULATION**

This simulation follows the method developed in prior research on time and income poverty, which is 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

---

<sup>9</sup> The average total transfers for individuals in adjusted consumption poor households receiving transfers is 221 thousand Won per month, compared to the average adjusted poverty line for such individuals of 1.5 million Won per month.

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, Occupation, and Employment Type**

The first step in assigning jobs to recipients is to determine what are the likeliest industry, occupation, and employment type for each of the recipients. This is done using a multinomial logit procedure. Industry, occupation, and employment type are regressed on age, age squared, sex, marital status, education and geographic region 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 are performed using a three-stage Heckit procedure (Berndt 1991, p. 627, separately for each combination of six age categories and sex. The first stage is a probit estimation of labor force participation:

$$lf_i = \alpha_1 + \beta X + \varepsilon_i \quad (1)$$

The vector of explanatory variables,  $X$ , comprises the number of children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, the individual's education, and the individual's spouse's age, education and labor force status. The regression is run on the universe of all eligible adults. The Mills ratio is calculated for all individuals using the results of the first stage regression:

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

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

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

$$\ln w_i = \alpha_2 + \gamma_2 Z + \theta_2 \lambda + \mu_i \quad (3)$$

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, occupation, employment type, spouse's labor force status, and finally,  $\lambda$ , 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, occupation, and employment type replaced for the latter by the assigned industries, occupations, and employment types from the first step.

The third stage is a regression of usual hours of paid work per week:

$$h_i = \alpha_3 + \gamma_3 Z + \omega \ln w_i + \theta_3 \lambda + \eta_i \quad (4)$$

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 the number of children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen 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.

### **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. The variables used to assess nearness of match are family type, marital status, spouse's labor force status, educational attainment and full-time part-time status, assigned industry, occupation, and employment type, the number of children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, and the two imputed variables, log of wage and hours worked. Industry and occupation are the most heavily

weighted variables with employment type as the next most heavily weighted variable. Next are imputed hours and wage, followed by family type and spouse's full-time/part-time status, then marital status and spouse's education and labor force status, and then the variables detailing the number 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 (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 adult 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, new hours caring for young children (since we will be reassigning child care hours contracted in the next stage), and new commuting hours to each of the adults, 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 family type, marital status, spouse's full-time/part-time status, number of adults, number of children, and the number of children aged less than one, one to two, three to five, six to twelve, and thirteen to seventeen in the household, household income, the income share of each individual,<sup>10</sup> 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. In this round of hot-decking, the number of children and number of adults in the household are weighted most

---

<sup>10</sup> Income share is included to reflect changes in bargaining power within the household and its impact on the distribution of household production work.

heavily of all the variables. Next most heavily weighted are family type and imputed usual hours of paid employment from the first stage. Then, the three variables detailing the number of young children in the household, household income, and income share followed by the remaining four employment-related variables assigned in the previous hot-decking step, and then the remaining variables detailing the number of older 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 income poverty, but we still need to adjust household childcare hours and total household production hours to reflect the new, higher household incomes and hours of paid work of recipient households.

### **Household Time Use and Child Care Services Reassignment**

In order to estimate the change in total household hours of household production and caring for young children as well as childcare services paid for and subsidized for recipient households we do a third hot-decking procedure, this time at the household level. In this stage we construct cells using the number of children and number of adults in the household and do all of the assignments within these cells. The matching variables are family type, age and educational achievement of the household head, assigned or actual employment type of the household head, the spouse of the household head's age, educational achievement and assigned or actual employment type, the number of non-working adults in the household, and the number of young children. In addition, we use the mother of the children in the household's full-time/part-time status (either actual or assigned), as well as the adjusted household income and imputed total household hours of household production and paid works hours. The family type variable was weighted most heavily, followed by the number of young children. The next most heavily weighted variables were the number of earners and non-working adults in the household, the imputed total household production and paid work hours and the mother's full-time/part-time status, followed by the imputed household income. Next were the household head's educational status and the head's spouse's employment type.

For each match we transferred household monthly privately purchased and publically subsidized child care hours, the total hours of household production and the total household hours of caring for young children. Finally we used the imputed or actual shares of household production from the second round of hot-decking to divide up the imputed household total hours of household production among individuals in the household. We then similarly divided up the

total hours of caring for young children among household members. The resulting data set can now be used to estimate time and income poverty as a result of the simulation. 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 the number of child care hours contracted. The assessment is therefore limited to comparing the latter qualities among sub-groups of donor and recipient records. Since the recipient and donor pools are not balanced in terms of underlying characteristics, there is no reason to think that the resulting distributions should be similar to the distributions in the donor pools. Nevertheless, lacking alternatives, we do compare them.

First we compare the compositions of 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, with the percentages representing the share of the female and male recipient and donor pools). We can see that among women, the members of the donor pool are somewhat evenly distributed by education and tend to be younger, while those in the recipient pool tend to be on the older side, and concentrated among high school graduates and those with some post-secondary education. Among men the distribution by age is quite similar for the donor and recipient pools, while recipients are slightly less educated overall. To some 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. However our method is tailored to make sure that we are matching individuals that are as similar as possible.

We can compare the industry, occupation, and employment type assigned in the employment simulation to the likeliest industry, occupation, and employment type calculated in the first step of our procedure. This comparison is presented in Tables 8, 9, and 10, respectively. As we can see, the assignment matched the likely industry in 95% of cases, while for both occupation and employment type the match rate was over 98%. Assessing the earnings imputation is less straightforward, for the reasons mentioned above. If the recipient pool has characteristics that are associated with lower earnings (as is the case), we would expect lower earnings not similar earnings among recipients compared to donors. We do however compare the assigned earnings to actual earnings by matching cell to check that the results are plausible.

Figure 4 displays the ratio of 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% from equivalence, a sweet spot. The worst ratio is that for women aged 45 to 54 with a college degree. Fortunately, this group represents only 2,100 of the 621 thousand recipients in the simulation. Generally, the more populated a cell with donors and recipients, the closer the results of the simulation are to the donor pool. 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, caring for young children, and commuting time, we again first compare the recipient and donor pools. Figure 6 shows the comparison by matching cell. The recipient pool has many fewer individuals in the 55 to 64 and 65 and older categories, since the earnings pool was restricted to households with adults who were not working and 55 years old. The donor pool is more evenly distributed, although still more heavily weighted towards younger individuals. Once again, the recipient pool includes relatively more individuals that are less educated than does the donor pool. This is again more pronounced for women, although both donor and recipient pools have a majority of women with high school diplomas or less. Figure 7 presents a comparison based on sex, number of children in the household and number of adults. Here we see that most males and females in both the donor and recipient pools are in households with no children and two or more adults. Large numbers are also found in households with one child and three or more adults and two children and two adults, but the portions are larger in the recipient than the donor pool. The relatively balanced nature of the demographic characteristics in the recipient and donor pools makes the following comparisons perhaps more meaningful.

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, at least in the case of the more populated subgroups. Among women with high school diplomas or less (comprising 75% of the women in the recipient pool), the average weekly hours are off by as much as 33%. This is slightly higher than we would like to see, but does not look unreasonable. For men, the averages are all slightly higher in the recipient pool and the medians are much higher, at least in percentage terms. However the denominator is



small in all cases, exaggerating the percentage variation, and where it is largest, the cell sizes are small. For example, the worst case is for the median weekly hours of household production for men aged 65 or older with college degrees. The recipient (there is only one record in the recipient pool in this cell) received 46 hours, while the median for the donor pool is 4 hours. Since this record represents only 484 individuals, this variation will not affect the overall results appreciably. Overall, the cases that are the furthest from equivalence are among elderly men and women, and these cells were, again, lightly populated. The comparison by sex and household composition is even better-looking. The worst case is for single males living alone who received a median of 18 weekly hours compared to the donor pool median of six hours. However, this represents 19 thousand recipients out of 1.3 million in the whole simulation.

The final assessment we do is of the household total household production and caring for young children and publically subsidized and privately paid for child care hours. This step happens at the household level. In this case the recipient and donor pools are divided up into cells based on the number of children and number of adults in the household. Figure 10 presents the comparison of the composition of the recipient and donor pools by these matching cells. In terms of the number of children, more donor households have none compared to the recipient households. However, since matching happens within these cells, there is no chance of a childless donor household being matched with a recipient household with 3 or more children. Figure 11 presents the ratio of the mean and median of the variables transferred in the third round of hot-decking in the recipient pool to those in the donor pool. In the most populated groups, especially the two-adult households, we see that most ratios are close to unity. There are some large differences in these groups, such as the two adult one child households, in which the recipients receive about 21 hours of publically subsidized care compared to the 7 hours for the donors. This is unsurprising given the fact that the donors include all households regardless of their income, while the recipients are all adjusted income poor households. None of the values for the transferred variables seem implausible.

#### **IV. CONCLUSION**

To the best of our ability to judge, the simulation looks like a reasonable approximation of the impact on individual adjusted income-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 do not account for loss of means-tested transfers. Nevertheless, the results should shed an interesting light on the impact of employment promotion on income poverty in Korea.

## 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](https://doi.org/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](https://doi.org/10.3233/JEM-2010-0332).
- Masterson, Thomas. 2010. "Quality of Match for Statistical Matches Used in the 1992 and 2007 LIMES Estimates for the United States." Levy Economics Institute Working Paper 618 (September). [http://www.levyinstitute.org/pubs/wp\\_618.pdf](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](http://www.levyinstitute.org/pubs/wp_727.pdf)
- \_\_\_\_\_. 2013. "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." Levy Economics Institute Working Paper 769 (July).  
[http://www.levyinstitute.org/pubs/wp\\_760.pdf](http://www.levyinstitute.org/pubs/wp_760.pdf)
- Zacharias, Ajit. 2011. "The Measurement of Time and Income Poverty." Levy Economics Institute Working Paper 690 (October). [http://www.levyinstitute.org/pubs/wp\\_690.pdf](http://www.levyinstitute.org/pubs/wp_690.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](http://www.levyinstitute.org/pubs/wp_568.pdf)

## Tables

*Table 1* Comparison of Actual and Predicted Presence in the Poverty Band for the KWPS 2009 Data for South Korea

<b>Poverty Band</b>	<b>Predicted Poverty Band</b>		<b>Total</b>
	<b>0</b>	<b>1</b>	
<b>0</b>	80.63	4.05	84.68
<b>1</b>	4.44	10.88	15.32
<b>Total</b>	85.07	14.93	100

Table 2 Alignment of Strata Variables, South Korea

	<b>KWPS 2009</b>	<b>KTUS 2009</b>	<b>Difference</b>
<b>Population</b>	43,219,236	43,297,959	-78,723
<b>Number of Children in Household</b>			
0	55.65%	58.52%	2.87%
1	17.22%	17.01%	-0.21%
2	23.04%	20.83%	-2.21%
3+	4.09%	3.64%	-0.45%
<b>Number of Adults (18yrs and Over) in Household</b>			
1	17.56%	20.79%	3.23%
2	53.31%	55.21%	1.90%
3+	29.13%	23.99%	-5.14%
<b>Household is Within the Poverty Band (y/n)</b>			
No	84.66%	80.40%	-4.26%
Yes	15.34%	19.60%	4.26%
<b>Presence of Non-Employed Adult in Household (y/n)</b>			
No	39.62%	46.71%	7.09%
Yes	60.38%	53.29%	-7.09%
<b>Household Income Category</b>			
<i>Less than 1,500,000 won</i>	24.12%	30.36%	6.24%
<i>1,500,000 to 2,499,999 won</i>	18.54%	21.00%	2.46%
<i>2,500,000 to 3,499,999 won</i>	18.31%	19.21%	0.90%
<i>3,500,000 to 4,999,999 won</i>	18.14%	17.08%	-1.06%
<i>5,000,000 won or more</i>	20.88%	12.35%	-8.53%
<b>Gender</b>			
Male	49.76%	49.40%	-0.36%
Female	50.24%	50.60%	0.36%
<b>Individual is Employed (y/n)</b>			
No	46.67%	43.78%	-2.89%
Yes	53.33%	56.22%	2.89%

Table 3 Distribution of Matched Records by Matching Round, South Korea

<b>Round</b>	<b>Matched Individuals</b>	<b>Percent</b>	<b>Cumulative Percentage</b>
1	30,023,820	69.5%	69.5%
2	3,046,900	7.0%	76.5%
3	431,694	1.0%	77.5%
4	2,836,124	6.6%	84.1%
5	368,664	0.9%	84.9%
6	279,794	0.6%	85.6%
7	118,028	0.3%	85.9%
8	224,546	0.5%	86.4%
9	1,287,745	3.0%	89.4%
10	147,355	0.3%	89.7%
11	503,805	1.2%	90.9%
12	318,167	0.7%	91.6%
13	506,900	1.2%	92.8%
14	1,301,080	3.0%	95.8%
15	234,521	0.5%	96.3%
16	597,816	1.4%	97.7%
17	263,293	0.6%	98.3%
18	728,951	1.7%	100.0%
<b>Total</b>	<b>43,219,202</b>	<b>100.0%</b>	

Table 4 Distribution of Weekly Hours of Household Production in KTUS 2009 and Matched File

	<b>p90/p10</b>	<b>p90/p50</b>	<b>p50/p10</b>	<b>p75/p25</b>	<b>p75/p50</b>	<b>p50/p25</b>	<b>Gini</b>
<b>KTUS 2009</b>	.	6.225	.	34.125	3.412	10.000	0.627
<b>MATCH</b>	.	6.125	.	27.300	3.412	8.000	0.625

Table 5 Comparison of Mean and Median Time Use Variables in Matched File to KTUS 2009

<b>Average</b>	<b>Core</b>	<b>Procurement</b>	<b>Care</b>	<b>Household Production</b>
<b>KTUS 2009</b>	9.97	1.08	3.47	14.51
<b>MATCH</b>	9.88	1.05	3.41	14.37
<b>Ratio</b>	99.10%	97.22%	98.27%	99.04%
<b>Median</b>	<b>Core</b>	<b>Procurement</b>	<b>Care</b>	<b>Household Production</b>
<b>KTUS 2009</b>	4.08	0.00	0.00	6.67
<b>MATCH</b>	4.08	0.00	0.00	6.67
<b>Ratio</b>	100.00%			100.00%

Table 6 Mean and Median Weekly Hours of Household Production by Strata Variable, KTUS 2009 and Matched File

**Mean Weekly Hours of Household Production**

	<b>KTUS 2009</b>	<b>MATCH</b>	<b>Ratio</b>			
<b>Core</b>	9.97	9.88	99.1%			
<b>Procurement</b>	1.08	1.05	97.2%			
<b>Care</b>	3.47	3.41	98.3%			
<b>Household Production</b>	14.51	14.37	99.0%			
<b>Number of Children</b>					<b>KTUS 2009</b>	<b>MATCH</b>
<b>0 children</b>	13.22	13.29	100.5%			
<b>1 child</b>	16.12	15.26	94.7%	<b>1/0</b>	1.22	1.15
<b>2 children</b>	15.74	15.72	99.9%	<b>2/0</b>	1.19	1.18
<b>3 or more children</b>	15.83	15.46	97.7%	<b>3+/0</b>	1.20	1.16
<b>Number of Adults</b>						
<b>1 adult</b>	14.07	14.41	102.4%			
<b>2 adults</b>	16.54	16.71	101.0%	<b>2/1</b>	1.18	1.16
<b>3 or more adults</b>	11.66	11.80	101.2%	<b>3+/1</b>	0.83	0.82
<b>Non-employed Adult in Household (y/n)</b>						
<b>No</b>	11.49	11.72	102.0%	<b>yes/no</b>	1.46	1.34
<b>Yes</b>	16.83	15.70	93.3%			
<b>Within Poverty Band (y/n)</b>						
<b>No</b>	14.07	14.03	99.7%	<b>yes/no</b>	1.22	1.20
<b>Yes</b>	17.17	16.77	97.7%			
<b>Household Income Category</b>						
<b>Less than 1,500,000 won</b>	17.00	17.10	100.6%			
<b>1,500,000 to 2,499,999 won</b>	15.73	15.44	98.2%	<b>2nd/1st</b>	0.93	0.90
<b>2,500,000 to 3,499,999 won</b>	14.78	15.46	104.6%	<b>3rd/1st</b>	0.94	1.00
<b>3,500,000 to 4,999,999 won</b>	12.50	12.94	103.5%	<b>4th/1st</b>	0.85	0.84
<b>5,000,000 won or more</b>	11.99	12.23	102.0%	<b>Top/1st</b>	0.96	0.95
<b>Employed (y/n)</b>						
<b>No</b>	19.31	18.42	95.4%	<b>yes/no</b>	0.56	0.59
<b>Yes</b>	10.77	10.83	100.6%			
<b>Sex</b>						
<b>Male</b>	23.74	23.14	97.5%	<b>Fem/Male</b>	0.21	0.24
<b>Female</b>	5.06	5.52	109.1%			

### Median Weekly Hours of Household Production

	KTUS 2009	MATCH	Ratio		KTUS 2009	MATCH
<b>Core</b>	4.08	4.08	100.0%			
<b>Procurement</b>	0.00	0.00				
<b>Care</b>	0.00	0.00				
<b>Household Production</b>	6.67	6.67	100.0%			
<b>Number of Children</b>						
<b>0 children</b>	7.58	7.58	100.0%			
<b>1 child</b>	7.00	6.42	91.7%	<b>1/0</b>	0.53	0.48
<b>2 children</b>	5.25	5.00	95.2%	<b>2/0</b>	0.40	0.38
<b>3 or more children</b>	4.50	4.17	92.7%	<b>3+/0</b>	0.34	0.31
<b>Number of Adults</b>						
<b>1 adult</b>	10.67	11.67	109.4%			
<b>2 adults</b>	8.00	7.67	95.9%	<b>2/1</b>	0.57	0.53
<b>3 or more adults</b>	4.08	4.67	114.5%	<b>3+/1</b>	0.29	0.32
<b>Non-employed Adult in Household (y/n)</b>						
<b>No</b>	6.00	6.42	107.0%	<b>yes/no</b>	1.17	1.09
<b>Yes</b>	7.00	7.00	100.0%			
<b>Within Poverty Band (y/n)</b>						
<b>No</b>	5.83	6.00	102.9%	<b>yes/no</b>	2.20	2.04
<b>Yes</b>	12.83	12.25	95.5%			
<b>Household Income Category</b>						
<b>Less than 1,500,000 won</b>	12.50	12.83	102.6%			
<b>1,500,000 to 2,499,999 won</b>	7.00	7.00	100.0%	<b>2nd/1st</b>	0.56	0.55
<b>2,500,000 to 3,499,999 won</b>	5.33	5.83	109.4%	<b>3rd/1st</b>	0.76	0.83
<b>3,500,000 to 4,999,999 won</b>	4.67	5.00	107.1%	<b>4th/1st</b>	0.88	0.86
<b>5,000,000 won or more</b>	4.67	5.25	112.4%	<b>Top/1st</b>	1.00	1.05
<b>Employed (y/n)</b>						
<b>No</b>	9.33	9.33	100.0%	<b>yes/no</b>	0.62	0.59
<b>Yes</b>	5.83	5.50	94.3%			
<b>Sex</b>						
<b>Male</b>	20.17	19.25	95.4%	<b>Fem/Male</b>	0.09	0.12
<b>Female</b>	1.75	2.33	133.1%			



*Table 7* Ratio of Matched to KTUS 2009 Average Hours of Household Production for the Reference Groups

<b>Number of Children</b>	<b>Number of Adults</b>		
	<b>1</b>	<b>2</b>	<b>3+</b>
<b>0</b>	111.3%	99.7%	110.4%
<b>1</b>	71.8%	93.8%	86.9%
<b>2</b>	87.3%	110.1%	94.4%
<b>3+</b>	62.0%	94.8%	88.2%

Table 8 Likely and Assigned Industries for Labor Market Simulation Recipients

Assigned Industry	Likely Industry				
	Agriculture, Forestry, Fishery	Manufacturing	Accommodation and Restaurant	All Others	Total
<b>Agriculture, Forestry, Fishery</b>	<b>9,386</b>	-	-	-	9,386
<b>Manufacturing</b>	993	<b>328,905</b>	-	344	330,242
<b>Construction</b>	-	4,054	-	-	4,054
<b>Whole/Retail Sale</b>	-	-	4,772	-	4,772
<b>Transportation</b>	-	2,087	-	-	2,087
<b>Accommodation and Restaurant</b>	-	-	<b>54,647</b>	10,322	64,969
<b>Government</b>	-	-	2,982	-	2,982
<b>All Others</b>	-	-	4,280	<b>209,297</b>	213,577
<b>Total</b>	10,379	335,046	66,681	219,963	632,069
<b>Percent Match</b>	<b>90.4%</b>	<b>98.2%</b>	<b>82.0%</b>	<b>95.2%</b>	<b>95.3%</b>

Table 9 Likely and Assigned Occupations for Labor Market Simulation Recipients

Assigned Occupation	Likely Occupation								
	Professional Occupations	Office and Administrative Support Occupations	Service Occupations	Sales Occupations	Farming, Fishing, and Forestry Occupations	Craft and Related Occupations	Machine Operation and Production Occupations	Manual Work Occupations	Total
Professional Occupations	126,142	-	-	-	-	-	-	-	126,142
Office and Administrative Support Occupations	-	61,024	-	-	-	-	-	-	61,024
Service Occupations	-	7,068	1,609	-	-	-	-	-	8,677
Sales Occupations	-	-	-	6,591	-	-	-	-	6,591
Farming, Fishing, and Forestry Occupations	-	-	-	-	9,386	-	-	-	9,386
Craft and Related Occupations	-	-	-	-	-	31,812	-	-	31,812
Machine Operation and Production Occupations	-	-	-	-	-	-	128,727	-	128,727
Manual Work Occupations	-	4,426	-	-	-	-	-	255,284	259,710
<b>Total</b>	126,142	72,518	1,609	6,591	9,386	31,812	128,727	255,284	632,069
<b>Percent Match</b>	<b>100.0%</b>	<b>84.2%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>98.2%</b>

Table 10 Likely and Assigned Employment Types for Labor Market Simulation Recipients

Assigned Class of Worker	Likely Class of Worker			
	Regular Employee	Irregular Employee	Employer and Self-Employed	Total
<b>Regular employee</b>	<b>366,575</b>	-	-	366,575
<b>Irregular employee</b>	13,016	<b>244,632</b>	-	257,648
<b>Employer and self-employed</b>	-	-	<b>7,846</b>	7,846
<b>Total</b>	379,591	244,632	7,846	632,069
<b>Percent Match</b>	<b>96.6%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>97.9%</b>

## Figures

Figure 1 Ratio of Mean HH Production by Category (Match/KTUS 2009)

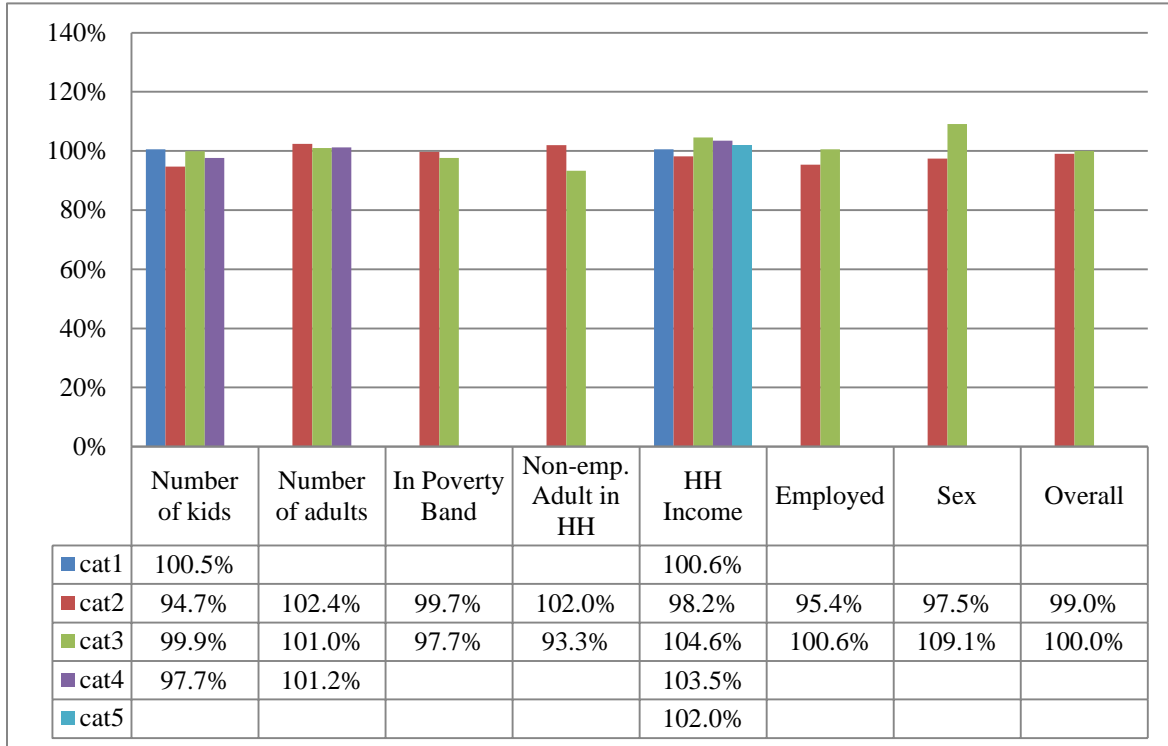


Figure 2 Household Production by Reference Groups, KTUS 2009 and Matched File

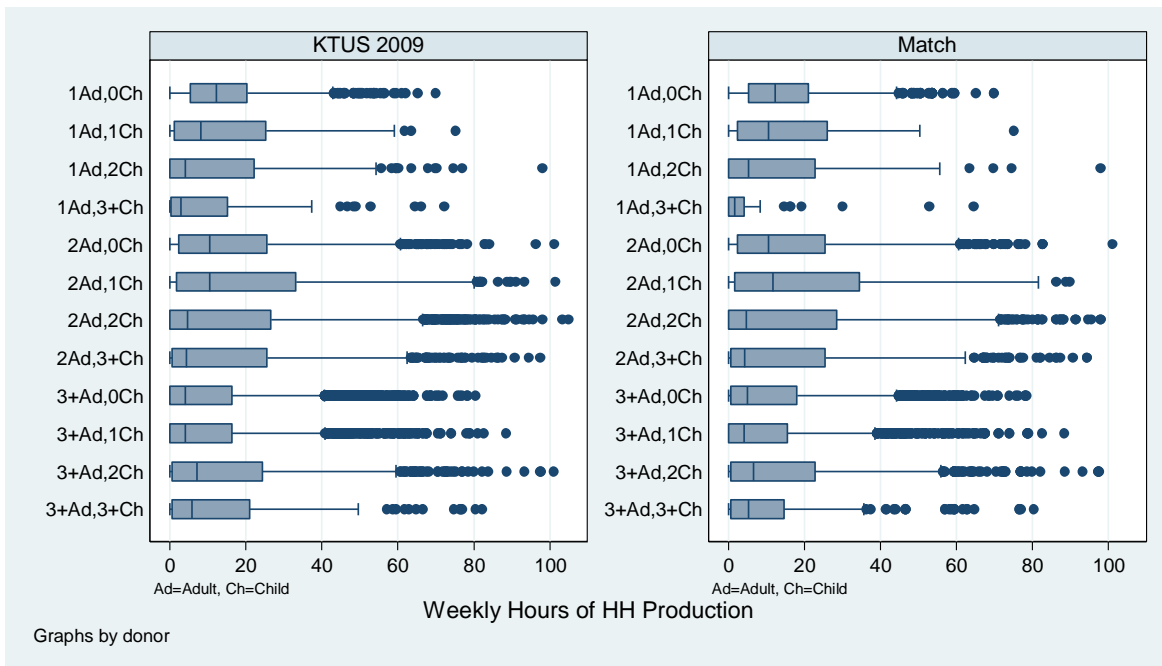


Figure 3 Donor and Recipient Pools for Labor Force Simulation by Sex, Age and Education

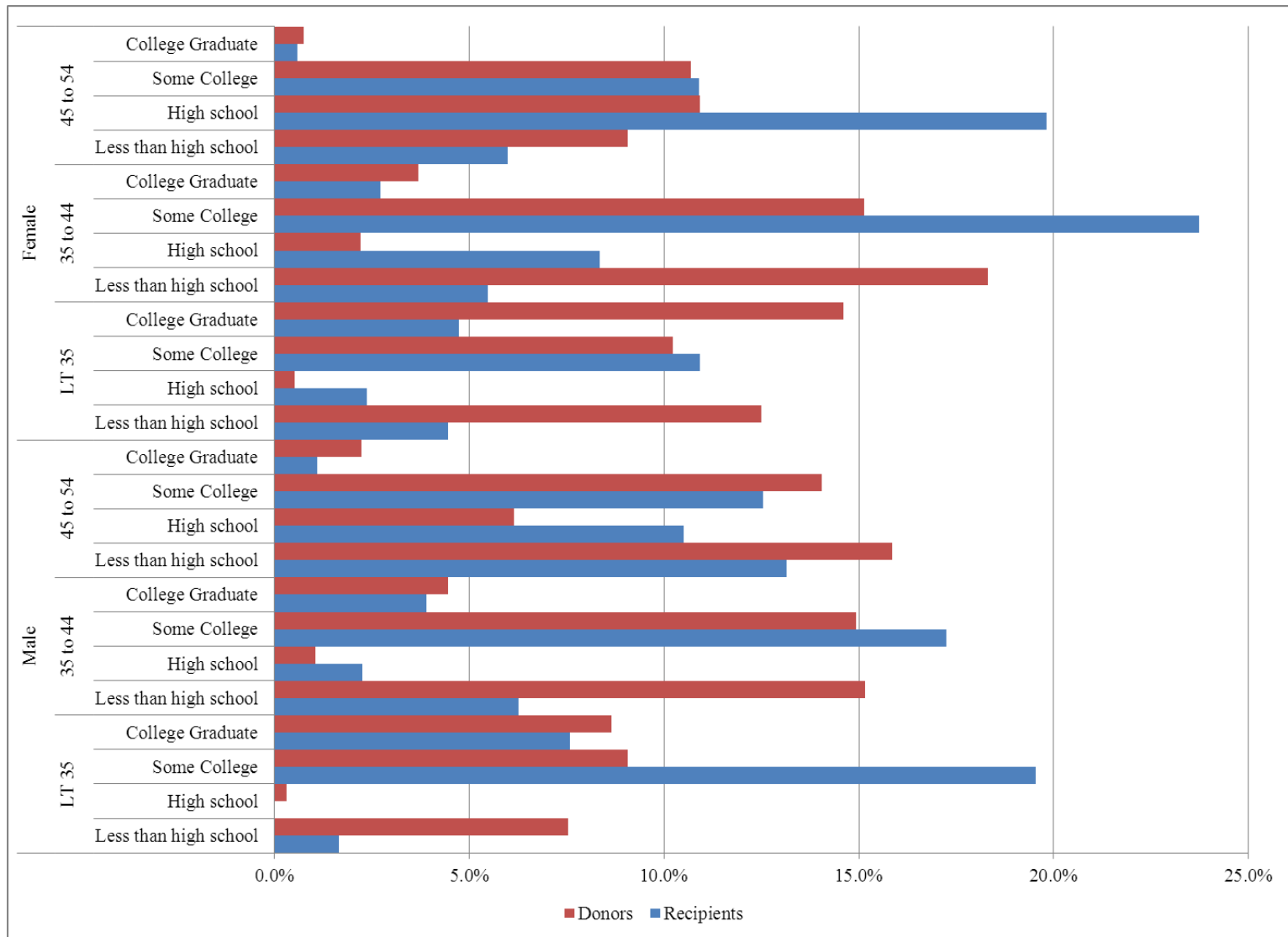


Figure 4 Ratios of Mean and Median Earned Income by Sex, Age and Education

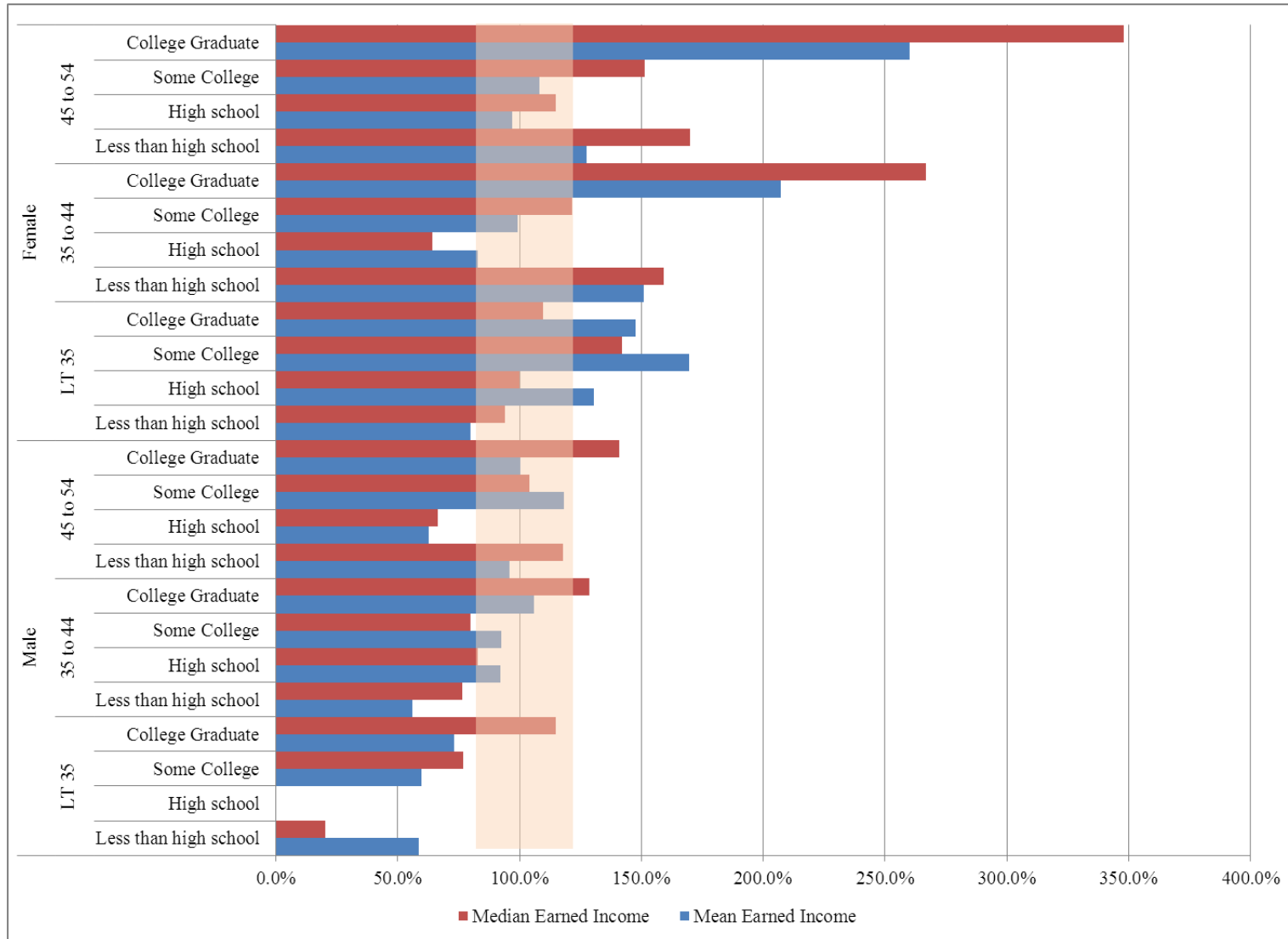


Figure 5 Ratios of Mean and Median Usual Hours of Work by Sex, Age and Education

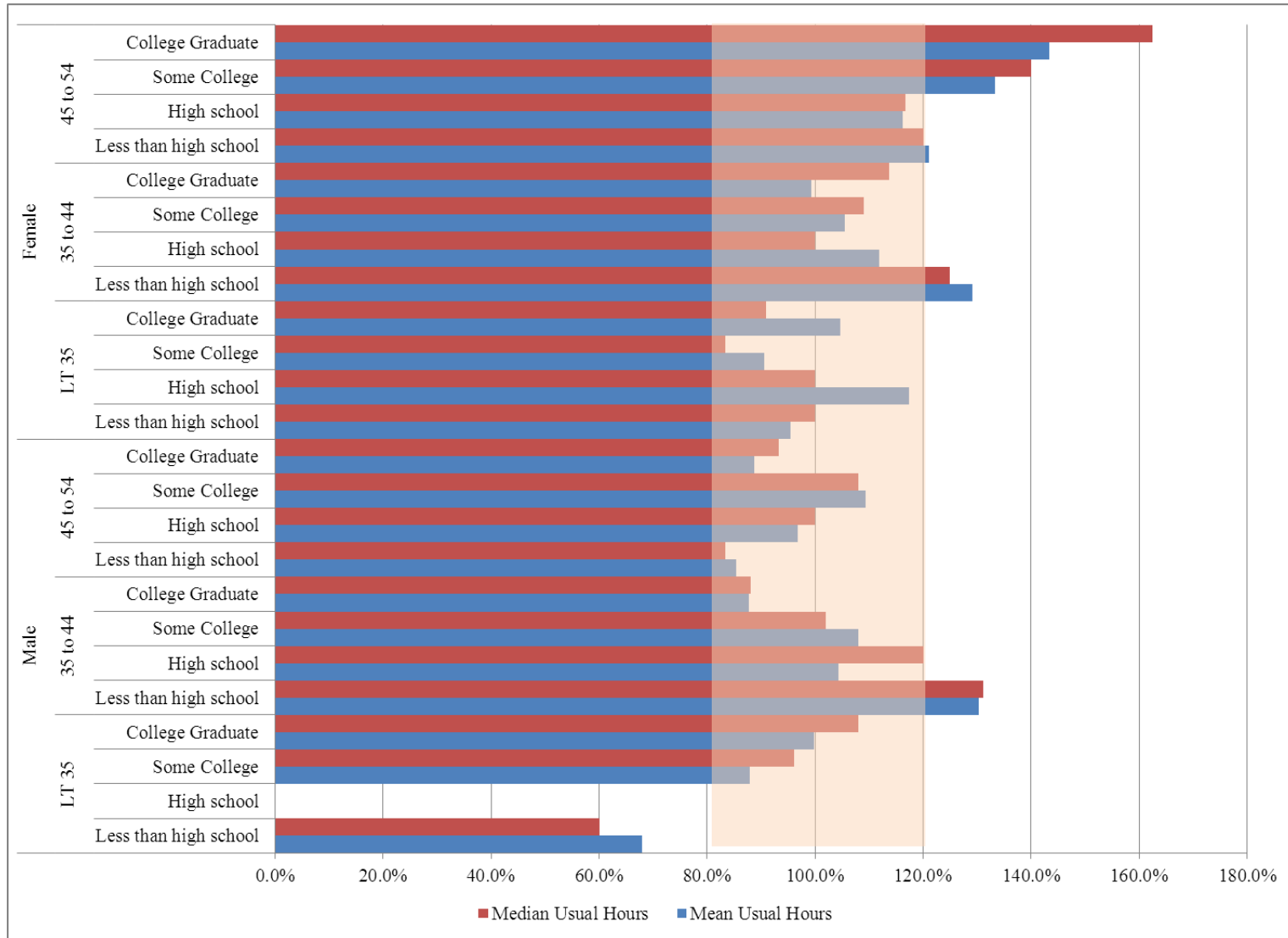




Figure 6 Donor and Recipient Pools for Time Use Simulation by Sex, Age and Education

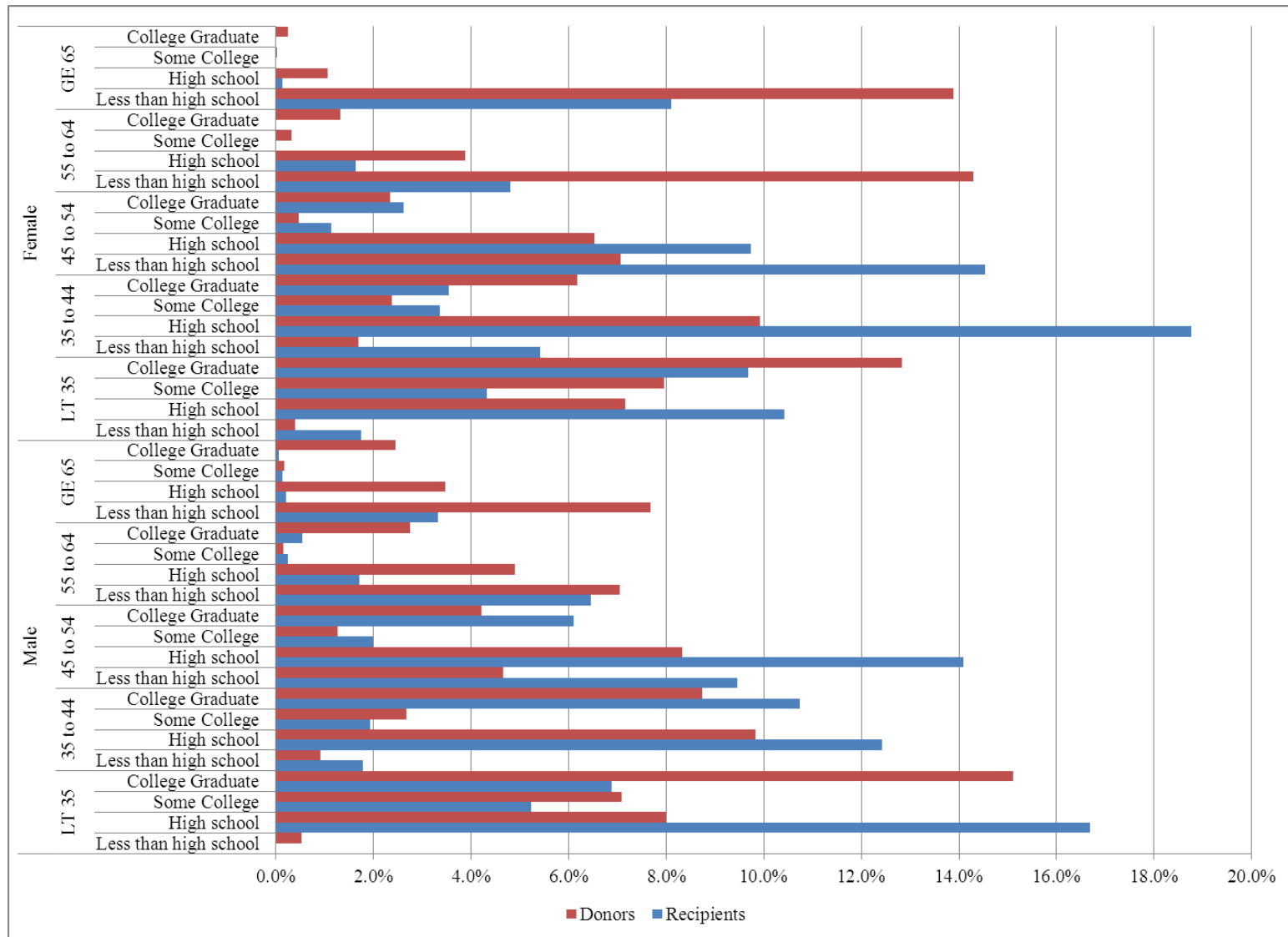


Figure 7 Donor and Recipient Pools for Time Use Simulation by Sex, Number of Adults and Number of Children

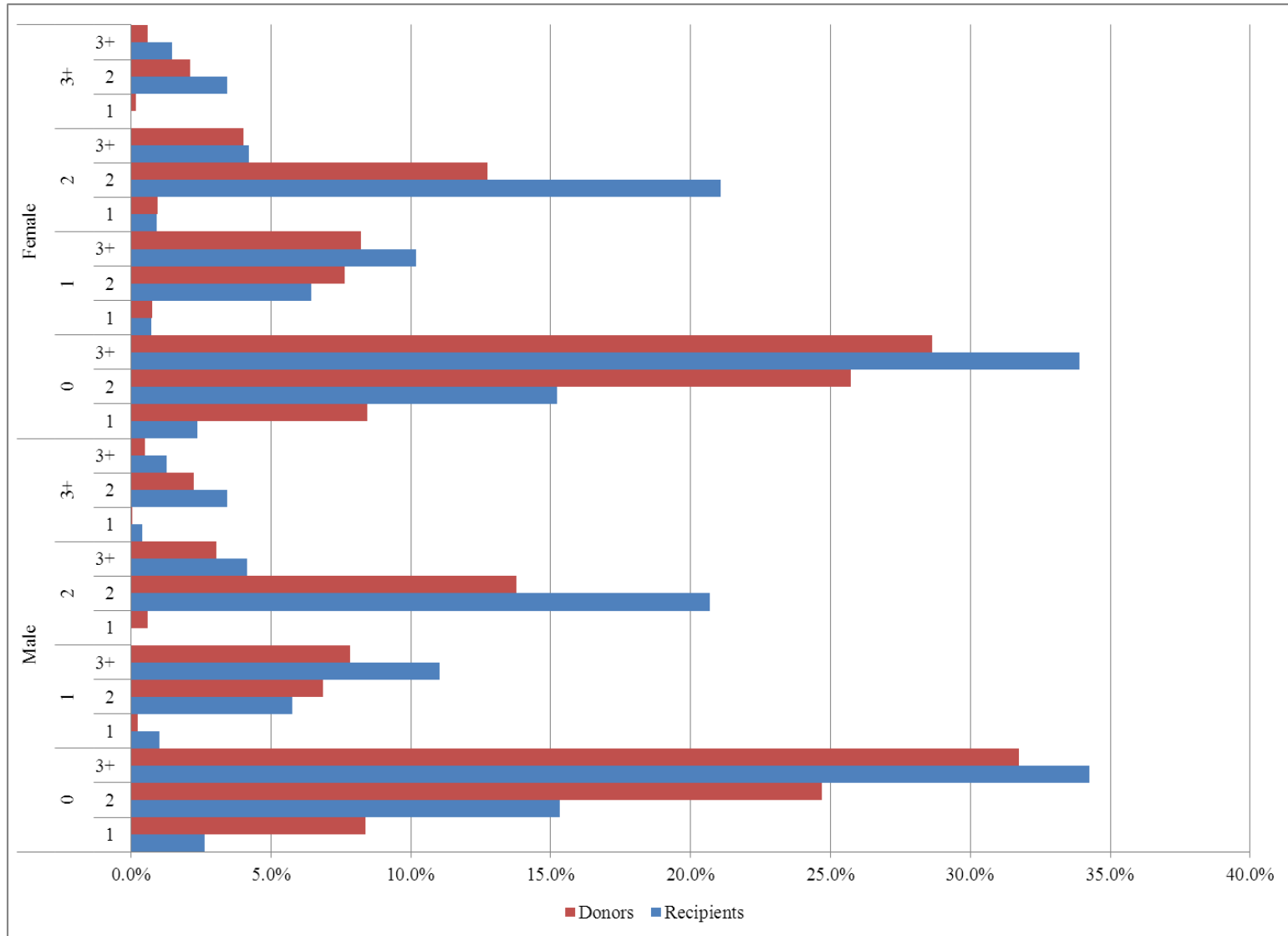


Figure 8 Ratios of Mean and Median Weekly Hours of Household Production by Sex, Age and Education

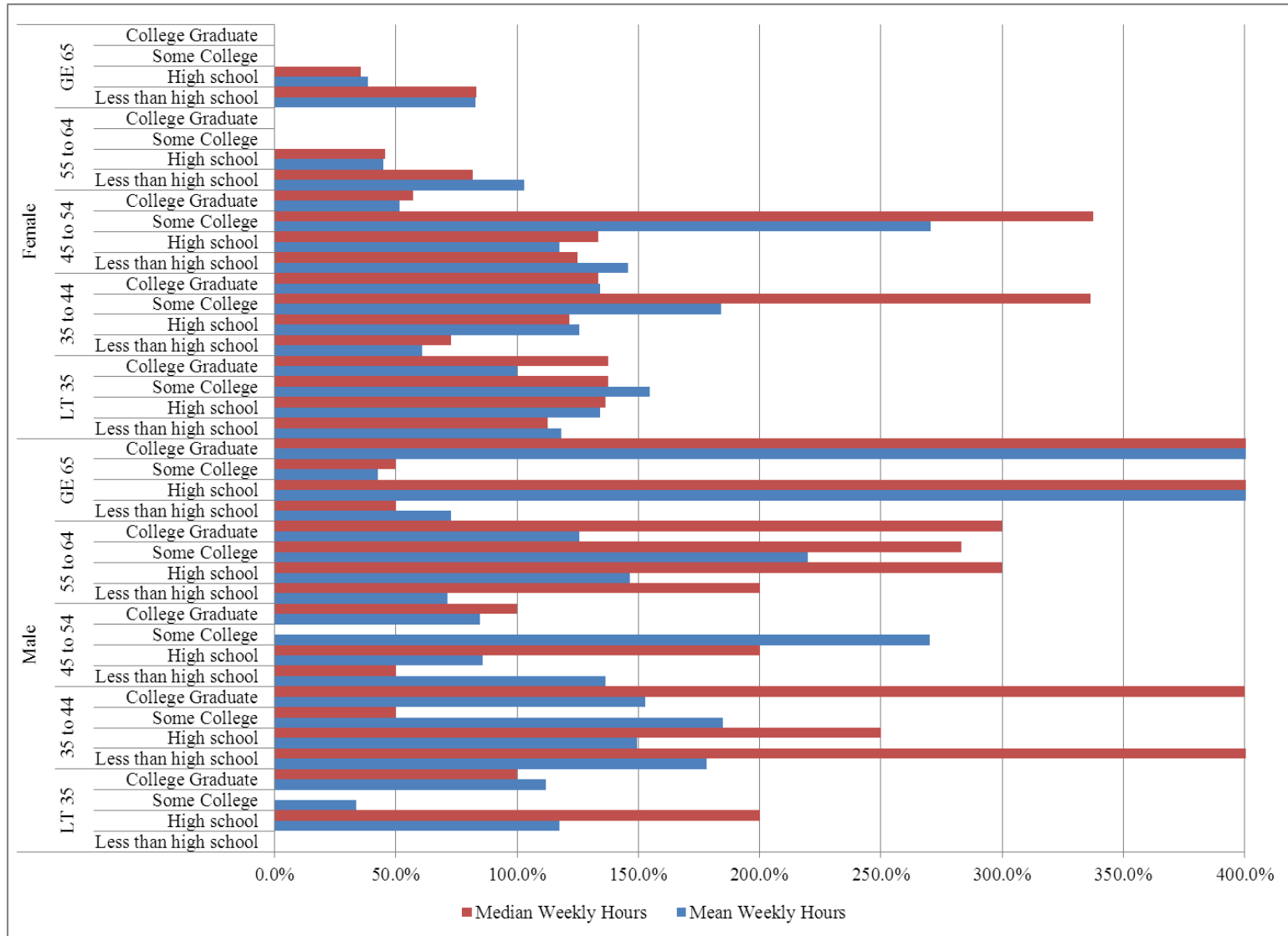


Figure 9 Ratios of Mean and Median Weekly Hours of Household Production by Sex, Number of Adults and Number of Children

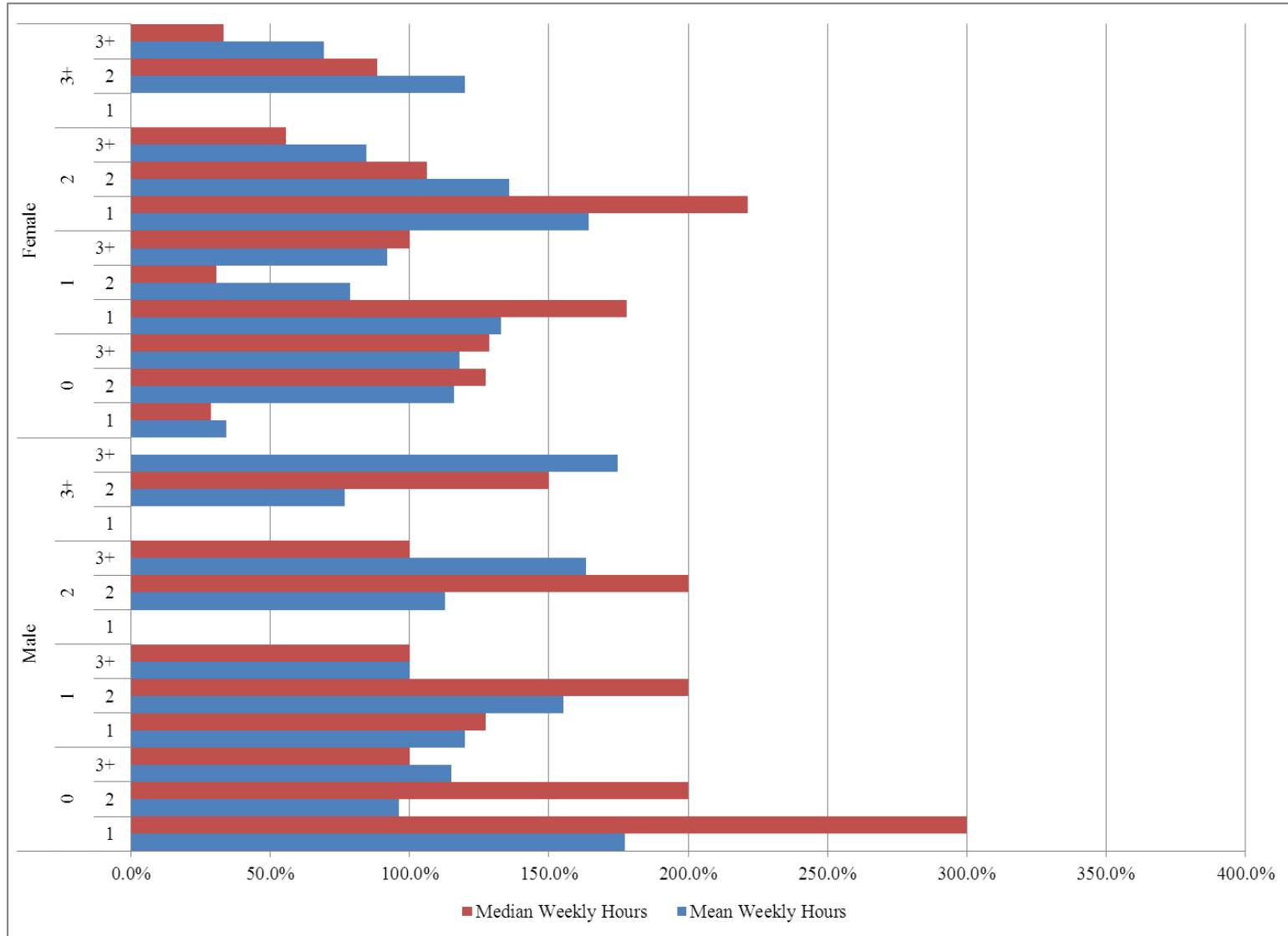


Figure 10 Donor and Recipient Pools for Childcare Hours Simulation by Number of Adults and Number of Children

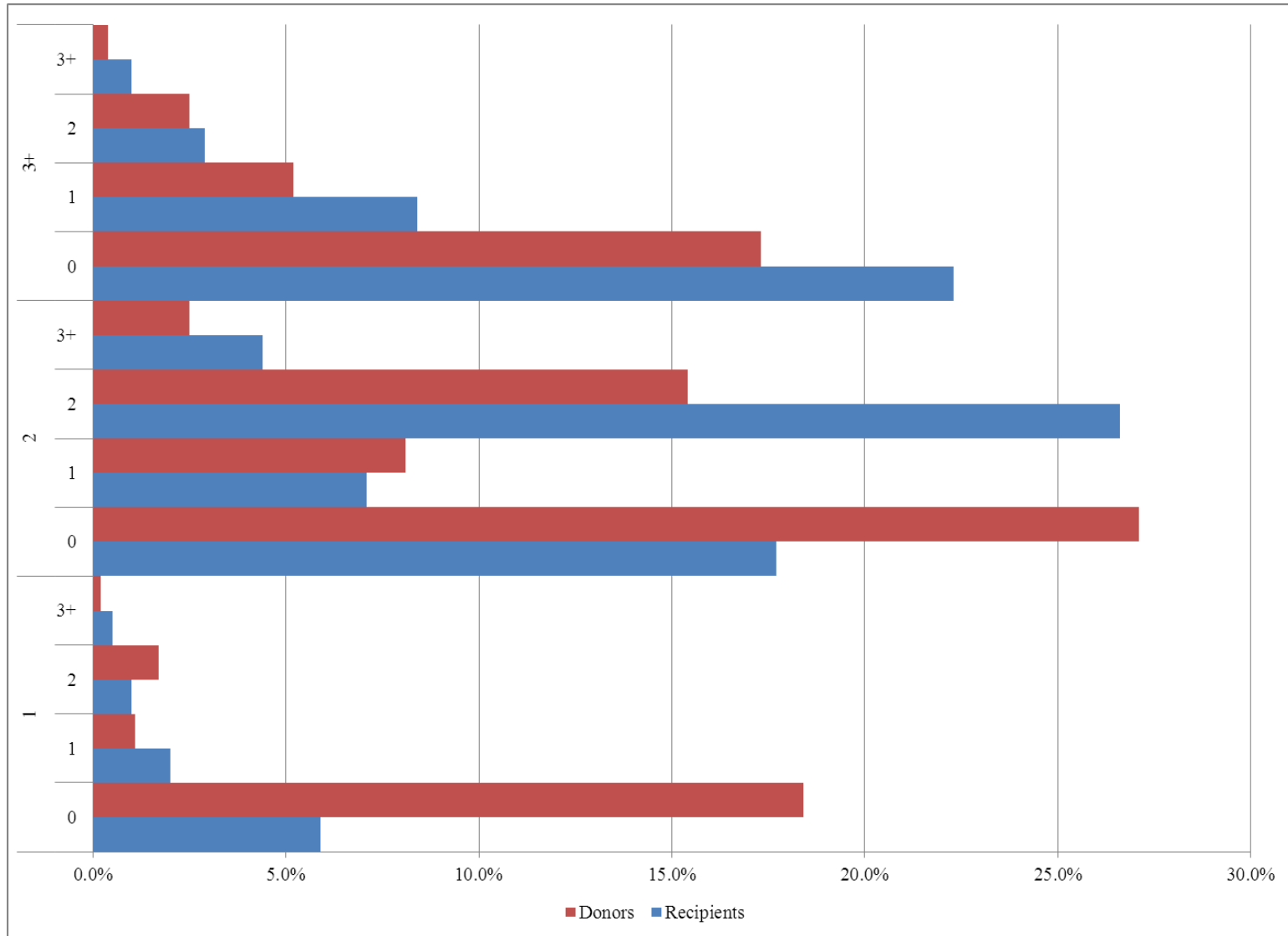


Figure 11 Ratios of Mean and Median Household Total Weekly Hours of Household Production, Privately Purchased and Publically Subsidized Child Care by Number of Children and Number of Adults

