Appendix A
Metadata for the 1995 CPS

1. The Structure of the CPS

As background, the CPS is composed of a core survey administered each month and a set of topical survey supplements that are administered in particular months in the calendar year (e.g., the displaced workers survey supplement is given in January, the annual income and demographic survey supplement is given in March). Households in the CPS are chosen by a clustered, stratified random sampling algorithm (see "Sampling" and "Source and Accuracy Statement" in the CPS documentation for details).

Importantly, the CPS has both a cross-sectional and a longitudinal structure. In particular, households are surveyed each month for four months (these are referred to as "months in the sample 1 through 4," respectively), then are out of the survey for eight months, then return to the survey for four additional months (these are referred to as "months in the sample 5 through 8," respectively). Consequently, the two sets of four monthly observations are one calendar year apart, affording a small longitudinal component to the CPS. As households exit the survey after the eighth month in the sample, they are refreshed with a new clustered, stratified random sample of replacement households, etc. Households that are in the sample in months 4 and 8 are called the "outgoing rotation groups" (ORG), since those in month 4 will be out of the survey for 8 more months, and those in month 8 will be out permanently.

2. CPS Wage Questions

The portion of the basic monthly survey concerning labor market activity that is asked in all months of the calendar year to all households in the survey is actually quite small. In fact, the CPS asks questions about wages on the current job(s) each month, but only to those households in the outgoing rotation groups (i.e., only those households that are in months in the sample 4 and 8).

For the purposes of the MATC CPS module, it does not matter which calendar month of the CPS is chosen for the module, since all calendar month CPS's have outgoing rotation groups with wage data. The CPS web site allows users to make extracts from the March, 1995 CPS, which is the most current CPS on the web site.

The dependent variable in the canonical wage equation in labor economics is the hourly wage, which is measured as the wage on the primary job that the household has at the time of the interview. Questions about this wage are asked only to those households in the March, 1995 CPS that are in the outgoing rotation groups (i.e., a subset, but a random subset, of all households in the March, 1995 CPS).

3. Construction of the Module Sample

The canonical wage equation in labor economics takes the form

$$\ln(w_i) = \mathbf{x}_i'\beta + \varepsilon_i$$

where w is the hourly wage, $\mathbf{x}$ is a vector of the determinants of wages (e.g., education, experience, union status, race, gender, etc.), β is a vector of parameters to be estimated, and ε is the error term. Labor economists are fond of this specification because when the dependent variable is the natural log of the hourly wage, the parameter with respect to education, for example, can be interpreted as the percentage change in wages for an additional unit (year) of educational attainment---or a rate of return to education.