# Effect of Introversion on Income
## Data Appendix

*Olivia Larsen (orl1), Robin Park (yp1), Jacob Watt-Morse (jww2)*

**Research Question**

How does the amount of *non work-related time spent alone* (indicative of trait introversion) affect *earnings* in the American labor market?

**Empirical Strategy**

In this project, we explore the effect of *time spent alone* on *wages*. The population we examine is Americans in the paid labor market in 2010. Sample data was taken from the American Time Use Survey (ATUS) from 2010, which has information on an individual's personal traits and use of time on a given calendar day.

We regress *time spent alone*, along with *presence and number of children*, *hours spent alone at work, total average hours spent working, presence of spouse or partner, gender, and age*, on the *log of weekly earnings*. The controls for presence and number of children are included in separate regressions.

**Codebook**

<u>Dependent Variable:</u>

Log of Hourly Wage (*LHOUR_WAGE)*, natural log of a respondent's reported wage

> The variable *TRERNHLY,* provided in the ATUS, represents the number of cents a respondent earns in an hour. For easier interpretation, we divide *TRERNHLY* by 100 to create *HOUR_WAGE*, which reflects the dollar amount that a respondent earn in an hour. Since we are more concerned with percent change (rather than dollar amount change) of wages, we use the natural log of *HOUR_WAGE* in our analysis, which we call *LHOUR_WAGE.*

> There are 6654 observations in the original dataset. We drop all respondents who reported zero as their earnings or did not report earnings at all (not a significant number), thoes who were not between the ages of 18-64, and those who worked inconsistent or no hours since the scope of our research limits our population to Americans who are in the paid labor market. The resulting number of observations is 3381.

> The mean of *LHOUR_WAGE*  is 2.637. The median is  2.639. The standard deviation is .522, the minimum reported value is 0, and the maximum is 4.605.
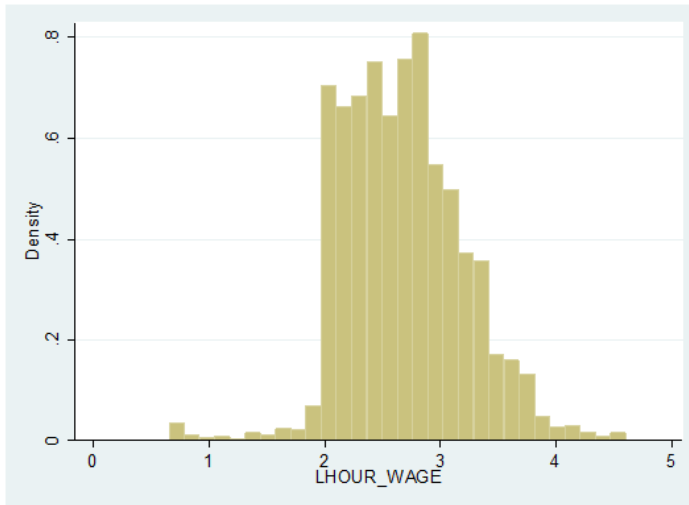
Independent Variable:

Hours per Day Spent Alone (*TRTALONE_HOUR)*, number of nonwork hours the respondent spent alone on the specified diary day.

> The variable *TRTALONE* is provided by the ATUS. We divide this by 60 to create *TRTALONE_HOUR*, which measures how many hours the respondent spent alone.

> There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

> The mean of *TRTALONE_HOUR* is 3.992. The median is 3.133. The standard deviation is 3.408, the minimum is 0, and the maximum is 21.
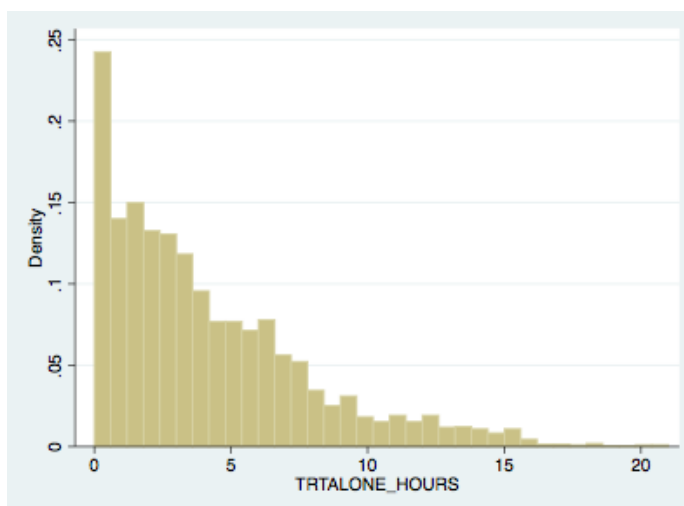
Control Variables:

Total Hours Usually Worked per Week (*TEHRUSLT*), number of hours a respondent usually worked per week.

The variable *TEHRUSLT* is provided by the ATUS.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *TEHRUSLT* is 37.291. The median is 40. The standard deviation is 11.567, the minimum is 0, and the maximum is 100.
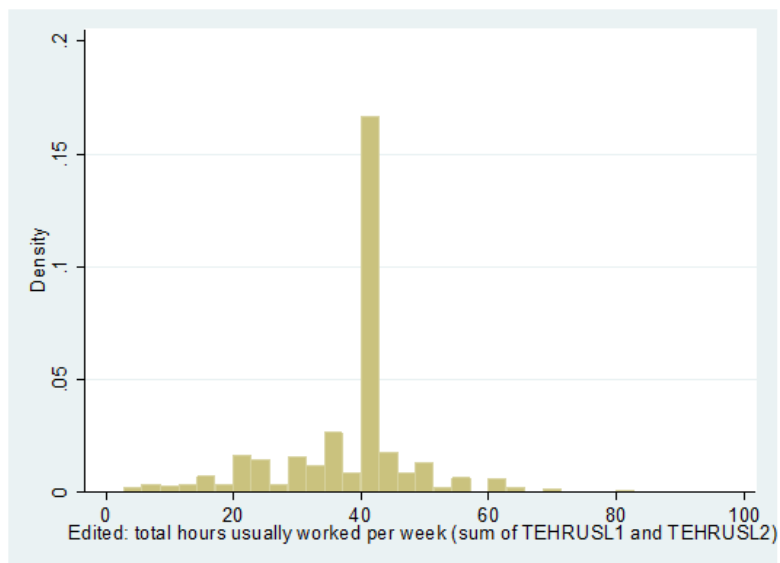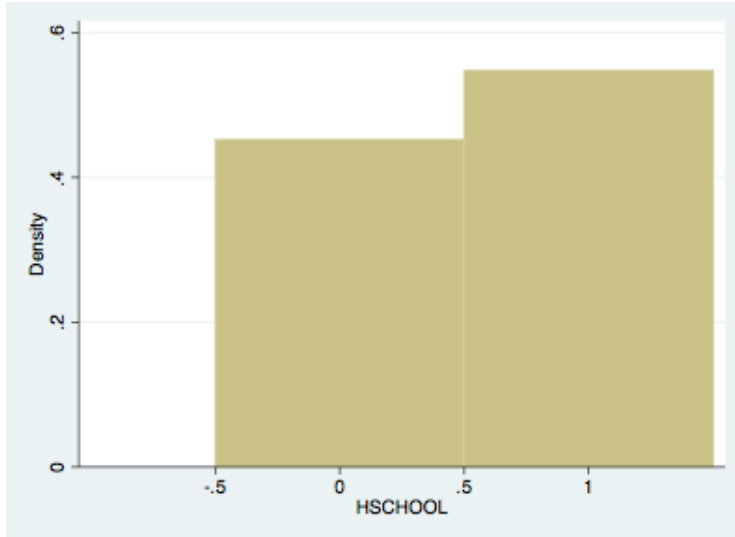


High School Dummy (*HSCHOOL*), dummy which indicates if a high school degree is the individual's highest level of educational attainment.

The variable *PEEDUCA* is provided by the ATUS that reports the highest level of education    the respondent has completed. *HSCHOOL* codes those whose highest level of educational attainment is a high school degree as 1 and those who did not complete high school or complete a higher degree as 0.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *HSCHOOL* is .582. The standard deviation is .493, the minimum is 0, and the maximum is 1.
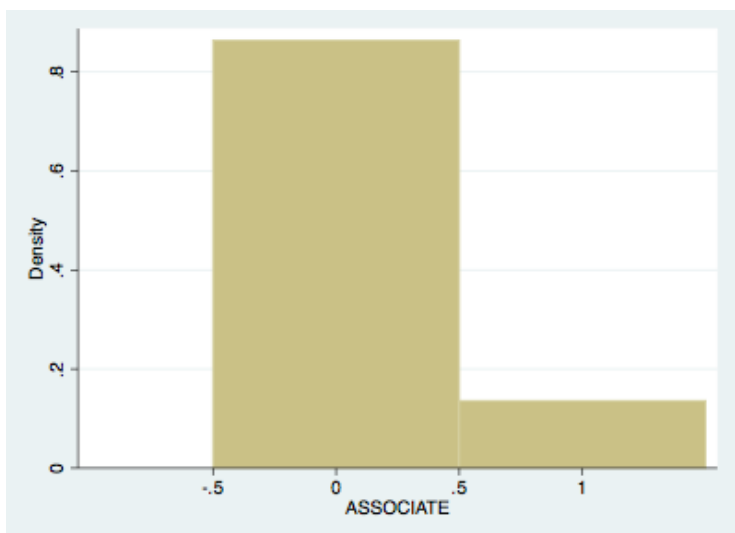
Associates Degree Dummy (*ASSOCIATE*), dummy which indicates if an associates degree was the highest level of educational attainment for a respondent.

The variable *PEEDUCA* is provided by the ATUS that reports the highest level of education the respondent has completed. *ASSOCIATE* codes those whose highest level of educational attainment is an associates degree as 1 and those who completed a lower or higher degree as 0.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *ASSOCIATE* is .108. The standard deviation is .310, the minimum is 0, and the maximum is 1.

Bachelors Degree Dummy (*BACHELOR*), dummy which indicates if a respondent received at least bachelors degree but could have also obtained a higher degree.

The variable *PEEDUCA* is provided by the ATUS that reports the highest level of education        the respondent has completed. *BACHELOR* codes those whose highest level of educational attainment is a bachelors degree as 1 and those who completed a lower or higher degree as 0.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *BACHELOR* is .196. The standard deviation is .397, the minimum is 0, and the maximum is 1.
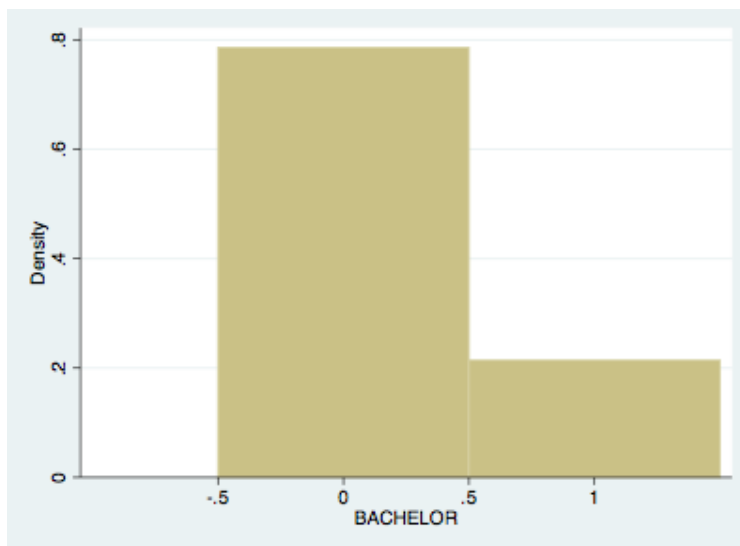


Male Dummy (*MALE*), dummy which indicates respondent's gender.

The variable *TESEX* is provided by the ATUS. *TESEX* reports gender with a male as 1 and female as 2. We converted this variable into a dummy, *MALE*, in which male is 1 and female is 0.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *MALE* is .503. The standard deviation is .5, the minimum is 0 and maximum 1.
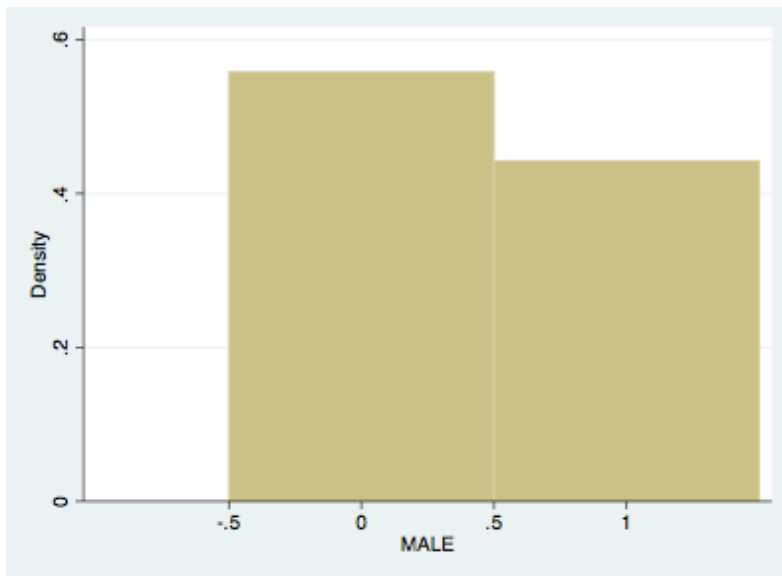
Age (*PRTAGE*), age of respondent.

*PRTAGE* is provided by the ATUS.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of age is 38.33. The median age was 40. The standard deviation of age 12.9385. The minimum age was 18 and the maximum was 64.
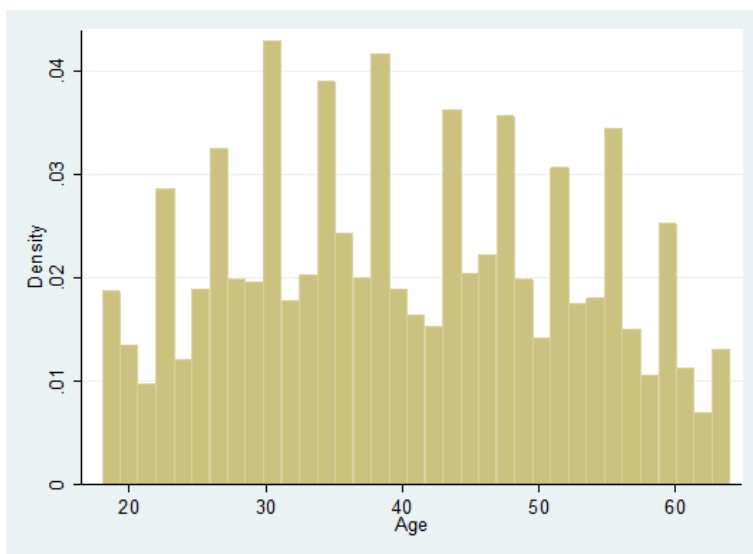
Age Squared (*AGESQ*), square of the age of respondents.

We derived *AGESQ* from *PRTAGE*, which is provided by the ATUS. We square *PRTAGE* to get *AGESQ*, which more accurately reflects the shape of the wage curve over an individual's career timeline.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of age squared is 1636.432. The median is 1600. The standard deviation is 1031.212. The minimum is 324 and the maximum is 4096.
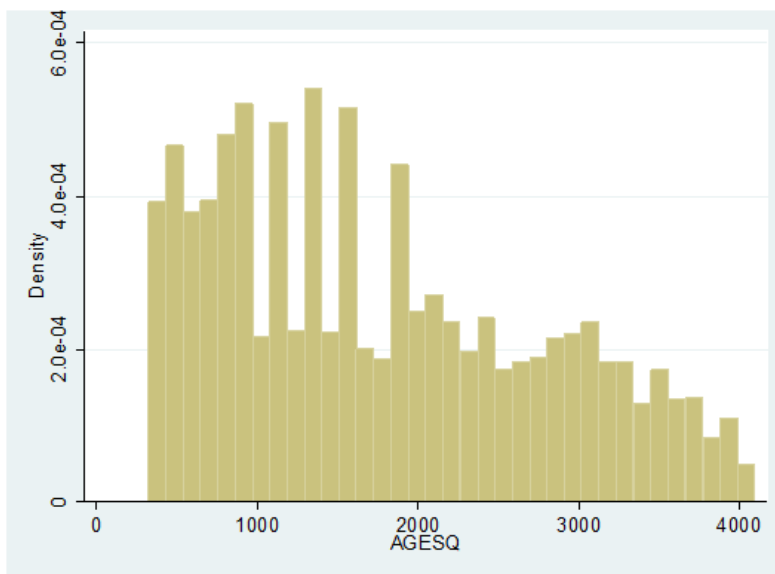


Time Spent Alone at Work (*TRTALONE_ONLYWORK*), daily time spent alone at work in minutes.

The *TRTALONE* and *TRTALONE_WK* variables are provided by the ATUS. We created *TRTALONE_ONLYWORK* by subtracting total time spent alone when not at work (*TRTALONE*) from total time spent alone (*TRTALONE_WK*).

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean in the sample was 56.117 minutes. The standard deviation was 152.248. The median was 0. The minimum amount of time spent alone was 0 minutes and the maximum was 990 minutes.
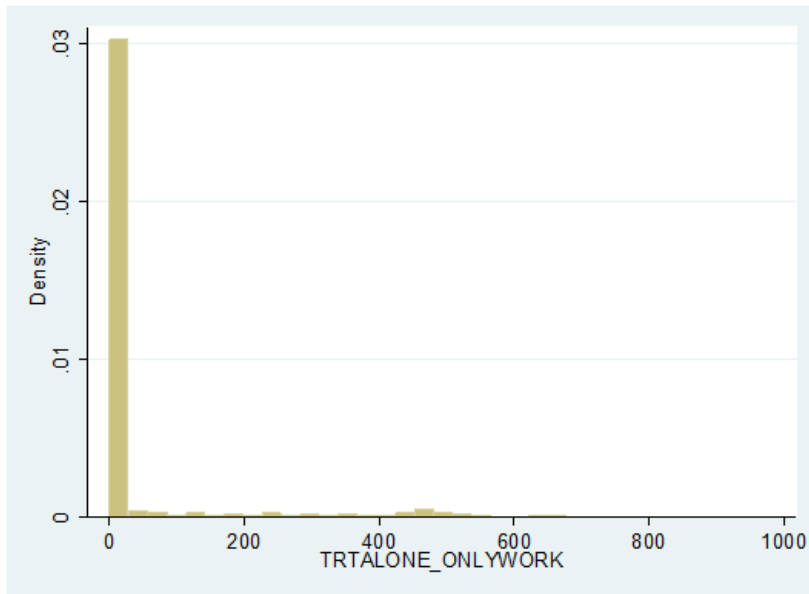
Presence of a Spouse (*TRSPPRES_DUM*), dummy which indicates if the respondent is currently living with a spouse or partner.

The variable *TRSPPRES* is provided by the ATUS that indicates the presence of a spouse as 1, a partner as 2 and neither as 3. *TRSPPRES_DUM* codes the presence of a spouse or partner as 1 and neither as 0.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *TRSPPRES_DUM* was .551. The standard deviation was .497 and the minimum is 0, and the maximum is 1.
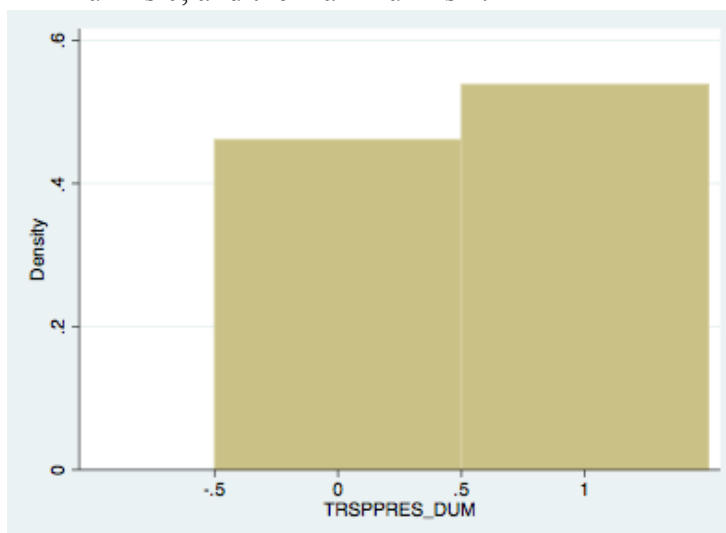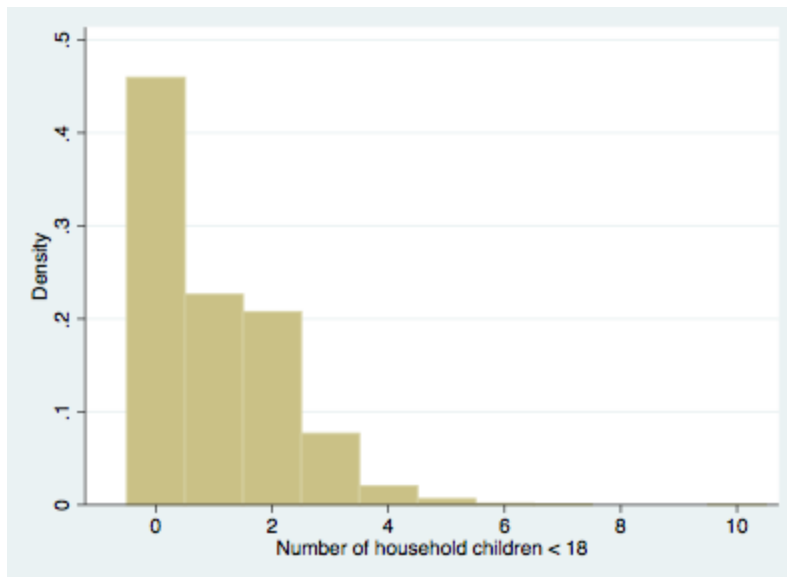
Number of Children in Household (*TRCHILDNUM*), number of children under 18 in respondent's household

The variable *TRCHILDNUM* is provided by the ATUS.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *TRCHILDNUM* is .844. The median is 0. The standard deviation is 1.166, the minimum is 0, and the maximum is 10.



Presence of Child (*TRHHCHILD_DUM*), dummy variable that indicates whether or not there are any children (defined as under age 18) in respondent's household.

The variable *TRHHCHILD* is provided by the ATUS. *TRHHCHILD* reports the presence of a household child as 2 and the lack of a household child as 1. We create the dummy variable *TRHHCHILD_DUM*, which codes the lack of a household child as 0 and the presence of a household child as 1.

There are 6654 observations in the original dataset, with no missing values. After we drop observations accordingly (for the same reasons noted for *LHOUR_WAGE*), the total number of observations is 3381.

The mean of *TRHHCHILD_DUM* is .448. The standard deviation is .497, the minimum is 0, and the maximum is 1.