**7-5-17 SAS code data sets**

* **gender\_inventor.csv**: This data set contains variables at the pooled TIN by tax year level (i.e. not cut by quartile) on the gender by inventor cuts of the data.
  + **payer\_tin\_w2\_max**: The employer TIN of the observation.
  + **tax\_yr**: The tax year of the observation.
  + **mean\_male\_inv\_wages**: This variable logs the mean wages of the male inventors working at the given firm in the given tax year.
  + **male\_invs**: This variable logs the number of male inventors working at the given firm in the given tax year.
  + **mean\_male\_noninv\_wages**: This variable logs the mean wages of the male non-inventors working at the given firm in the given tax year.
  + **male\_noninvs**: This variable logs the number of male non-inventors working at the given firm in the given tax year.
  + **mean\_female\_inv\_wages**: This variable logs the mean wages of the female inventors working at the given firm in the given tax year.
  + **female\_invs**: This variable logs the number of female inventors working at the given firm in the given tax year.
  + **mean\_female\_noninv\_wages**: This variable logs the mean wages of the female non-inventors working at the given firm in the given tax year.
  + **female\_noninvs**: This variable logs the number of female non-inventors working at the given firm in the given tax year.
* **quartile\_gender\_inventor.csv*:*** This data set contains variables at the TIN by tax year level by wage quartile level on the gender, inventor, and gender by inventor cuts of the data.
  + **payer\_tin\_w2\_max**: The employer TIN of the observation.
  + **tax\_yr:** The tax year of the observation.
  + **quartile**: The within-firm wage quartile of the observation in the given tax year.
  + **mean\_male\_q\_wages**: This variable logs the mean wages of males working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_males**: This variable logs the number of males working at the given firm in the given wage quartile in the given tax year.
  + **mean\_female\_q\_wage**s: This variable logs the mean wages of females working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_females**: This variable logs the number of females working at the given firm in the given wage quartile in the given tax year.
  + **mean\_inv\_q\_wages**: This variable logs the mean wages of inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_invs**: This variable logs the number of inventors working at the given firm in the given wage quartile in the given tax year.
  + **mean\_noninv\_q\_wages**: This variable logs the mean wages of non-inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_noninvs**: This variable logs the number of non-inventors working at the given firm in the given wage quartile in the given tax year.
  + **mean\_male\_inv\_q\_wages**: This variable logs the mean wages of the male inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_male\_invs**: This variable logs the number of male inventors working at the given firm in the given wage quartile in the given tax year.
  + **mean\_male\_noninv\_q\_wages**: This variable logs the mean wages of the male non-inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_male\_noninvs**: This variable logs the number of male non-inventors working at the given firm in the given wage quartile in the given tax year.
  + **mean\_female\_inv\_q\_wages**: This variable logs the mean wages of the female inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_female\_invs**: This variable logs the number of female inventors working at the given firm in the given wage quartile in the given tax year.
  + **mean\_female\_noninv\_q\_wages**: This variable logs the mean wages of the female non-inventors working at the given firm in the given wage quartile in the given tax year.
  + **quartile\_female\_noninvs**: This variable logs the number of female non-inventors working at the given firm in the given wage quartile in the given tax year.
* **mean\_quartile\_stayer\_currwages.csv**: This data set contains variables at the TIN by tax year level by wage quartile level on the wages for stayer wages in the current year and the application year computed by the current year’s quartiles.
* To get the composition adjusted values, simply subtract mean\_styrXXXX\_q\_oldwages from mean\_styrXXXX\_q\_currwages
* Note that the interpretability of all of the variables in this data set is flipped from the usual stayer values. While the usual stayer values would merge each tax year observation onto the application year and select the variable corresponding to the decision year, you should merge each tax year observation onto the decision year and select variables by the application year for this data set.
  + **payer\_tin\_w2\_max**: The employer TIN of the observation.
  + **tax\_yr:** The tax year of the observation.
  + **quartile**: The within-firm wage quartile of the observation in the given tax year.
  + **mean\_styrXXXX\_q\_currwages:** This variable measures the mean current wages of year XXXX stayers for a given firm in the given wage quartile in the given tax year. For example, say we are looking at the value of the 2005 version of this variable for the TIN 1000, tax year 2008, and quartile 3 observation. The value would measure the mean 2008 wages for stayers from 2005 who are in the third wage quartile in 2008 at firm 1000.
  + **mean\_styrXXXX\_q\_oldwages**: This variable measures the mean XXXX wages of year XXXX stayers for a given firm in the given wage quartile in the given tax year. For example, say we are looking at the value of the 2005 version of this variable for the TIN 1000, tax year 2008, and quartile 3 observation. The value would measure the mean 2005 wages for stayers from 2005 who are in the third wage quartile in 2008 at firm 1000.
* **mean\_entrant3s.csv**: This data set contains variables at the pooled TIN by tax year level (i.e. not cut by quartile) on the entrants within the past three years to the firm. As such, it contains data from 2003 to 2014 (as a full 3-year panel was not able to be constructed for 2000-2003).
  + **payer\_tin\_w2\_max**: The employer TIN of the observation.
  + **tax\_yr**: The tax year of the observation.
  + **mean\_entrant3\_wages**: This variable measures the mean wages of people who entered into the given firm within the past three years in the given tax year. For example, say we are looking at the value for the TIN 1000 and tax year 2008 observation. The value would measure the mean 2008 wages for people who were working at different firms or not working in either 2005, 2006, or 2007, and who are now working at firm 1000 in 2008.
  + **entrant3s**: This variable measures the number of people who entered into the given firm within the past three years in the given tax year.
* **quartile\_mean\_entrant3s.csv**: This data set contains variables at the TIN by tax year by wage quartile level on the entrants within the past three years to the firm. As such, it contains data from 2003 to 2014 (as a full 3-year panel was not able to be constructed for 2000-2003).
  + **payer\_tin\_w2\_max**: The employer TIN of the observation.
  + **tax\_yr**: The tax year of the observation.
  + **quartile**: The within-firm wage quartile of the observation in the given tax year.
  + **mean\_entrant3\_q\_wages**: This variable measures the mean wages of people who entered into the given firm within the past three years in the given wage quartile in the given tax year. For example, say we are looking at the value for the TIN 1000, tax year 2008, and quartile 3 observation. The value would measure the mean 2008 wages for people who were working at different firms or not working in either 2005, 2006, or 2007, and who are now working at firm 1000 in 2008 in the third wage quartile.
  + **quartile\_entrant3s**: This variable measures the number of people who entered into the given firm within the past three years in the given wage quartile in the given tax year.