**README for the Data Appendix to: “A Grand Gender Convergence: Its Last Chapter,” Claudia Goldin**

This README describes the various folders and files in the Zip File and maps them into the Figures and Tables in the paper. In most cases a STATA \*.do file is given as well as an XLS workbook containing the results. If the data are in the public domain (e.g., CPS, Census, ACS), the \*.do file states the data source. If the data must be obtained from an institution (e.g., University of Michigan Law School Alumni Research Dataset), that is listed as well. The only raw data contained in the Zip File are for the O\*Net computation.

**Figure 1**: See CohortWageGaps folder.

Figure 1: Relative Earnings of (Full-Time, Full-Year) College Graduate Men and Women for Synthetic Cohorts: Born 1923 to 1978

Part A: No Controls, Part B: With controls for work time and education

gender\_wage\_gap\_full\_census\_XXX, where XXX = year of census or ACS (e.g., 1980) contains the results of the regression. There are six of these files, each of which contains the \*.do file and the output.

WageGapCohorts092713.xls is a spreadsheet that contains the coefficients on the age group × female from the six files and creates the graphs. It should be noted that there is a small amount of interpolation using a simple linear method.

**Figure 2**: See Occupations folder.

Figure 2: Gender Pay Gaps by Occupation: 2009 to 2011

Part A: Full-time, full-year for the approximately 95 highest (male) income occupations, Part B: Full-time, full-year, college graduates (four years of college) for the approximately 95 highest (male) income occupations, Part C: Full-time, full-year, less than 45 years old for the approximately 95 highest (male) income occupations

analysis\_2011\_v2.log, contains the \*.do file and output for all occupations

AllOccsWageGaps.xls, contains results of the various regressions for different samples (e.g., BA, Full-time, Full-year; All workers). “Full-Educ-Time” contains the graph of the calculated gender pay gap against mean male income by occupation, showing no relationship. See the following worksheets for each part in the Figure: Part A: “Full-Educ-Time,” Part B: “Full-BA-Educ-Time.” Part C is contained in ByAge.xls, worksheet “Full\_Younger.” Note that the tabs that are used are colored; those that are not use different versions of the sample restrictions to test robustness.

AllOccsWageGaps.xls also contains various calculations, noted in this README. For example, the calculation regarding the portion of the gender wage gap due to differences in earnings within each occupation versus the distribution of individuals by occupation. The worksheet “Calcs” contains the calculation for the fraction of all college graduates included in the approximately 95 categorized occupations. The worksheet “AverageCoeffs” contains a computation of the average of the coefficients for each of the separate occupational groups, weighted by the numbers in each of the occupations.

**Figure 3**: See Occupations folder.

Figure 3: Relationship between the Elasticity of Earnings with Respect to Hours and the Gender Earnings Gap

AllOccsAgeGaps.xls, see worksheet “FullBA,EducTime plus Hours”

hrs\_intx.do

Regression-FullBA-EducTime-hrs.docx

[**Figure 4**: The Framework. No data are used.]

**Figure 5**: See ONet folder.

Figure 5: O\*Net Characteristics and the Residual College Gender Pay Gap by Occupation

ONet\_characteristics.xls, contains the O\*Net data

ONet\_characteristics.do

Readme\_ONet.docx, explains how the O\*Net data were obtained

OccCharacteristicOnet.xls contains worksheet “GG Graphs ONet” with the graph

OccCharacMerge.dta has the data for the regression line

**Figure 6**: See Law folder.

Figure 6: Hours, Earnings, and Characteristics of Law Occupations Fifteen Years after the JD

umls\_clean.do, creates the analysis longitudinal data file from asrdl.dta provided as the University of Michigan Alumni Law School Research Dataset.

lawyervars.do, labels the variables

CGlawyer.do is the \*.do file for Figure 5 (at the end) and lots of other regressions and tabs

CGlawyerLog.log contains the data for Figure 5

**Table 1**: See Occupations folder.

Table 1: Residual Gender Differences in Earnings and the Role of Occupation

WageRegswithOccs.log, contains the \*.do file and output

**Table 2**: See ONet folder.

ONet\_characteristics.xls

ONet\_characteristics.do

Readme\_ONet.docx

**Table 3**: See Law folder.

umls\_clean.do, creates the analysis longitudinal data file from asrdl.dta provided as the University of Michigan Alumni Law School Research Dataset.

lawyervars.do, labels the variables

CGlawyer.do is the \*.do file for cols. (1) to (8) and contains lots of other regressions and tabs as well

CGlawyerLog.log contains the regressions for cols. (1) to (8)