Example 4a: General Linear Models with Multiple Fixed Effects of Multiple Predictors Simultaneously *(complete syntax, data, and output available for SAS, STATA, and R electronically)*

The data for this example were selected from the 2012 General Social Survey dataset (and were also used for examples 1, 2, and 3). Building on the results of Example 3 (summarized below), this example will examine the unique effects of three-category working class, linear and quadratic slopes for years of age, and three piecewise slopes (i.e., linear splines) for years of education in predicting annual income. It will also demonstrate how the results from hierarchical (stepwise) regression can be obtained from a single model using multivariate Wald F-tests instead. See the syntax and output online for how to compute effect sizes per slope using their *t* statistics.

SAS Syntax for Importing and Preparing Data for Analysis:

```
* Paste in the folder address where "GSS_Example.xlsx" is saved after = before ;
%LET filesave= \\Client\C:\Dropbox\22SP_PSQF6243\PSQF6243_Example4a;

* IMPORT GSS_Example.xlsx data using filesave reference and exact file name;
* from the Excel workbook in DATAFILE= location from SHEET= ;
* New SAS file is in "work" library place with name "Example4a";
* "GETNAMES" reads in the first row as variable names;
* DBMS=XLSX (can also use EXCEL or XLS for .xls files);
PROC IMPORT DATAFILE="&filesave.\GSS_Example.xlsx"
            OUT=work.Example4a DBMS=XLSX REPLACE;
     SHEET="GSS_Example";
     GETNAMES=YES;
RUN;

* All data transformations must happen inside a DATA+SET combo to know where to use them;
DATA work.Example4a; SET work.Example4a;
* Create and label predictor variables for model 1 (same as in Example 2);
  * Linear predictor for education centered so that 0 is meaningful;
    educ12=educ-12;
    LABEL educ12= "educ12: Education (0=12 years)";
  * Recode binary marry predictor so that 0 is meaningful;
    marry01=.; * Create new empty variable, then recode;
    IF marry=1 THEN marry01=0;
    IF marry=2 THEN marry01=1;
    LABEL marry01= "marry01: 0=unmarried, 1=married";
* Create and label predictor variables for model 2 (same as in Example 3);
  * 2 Indicator-dummy-coded binary predictors for workclass;
    LvM=.; LvU=.; * Make new empty variables;
    IF workclass=1 THEN DO; LvM=0; LvU=0; END; * Replace each for lower;
    IF workclass=2 THEN DO; LvM=1; LvU=0; END; * Replace each for middle;
    IF workclass=3 THEN DO; LvM=0; LvU=1; END; * Replace each for upper;
    LABEL LvM="LvM: Low=0 vs Mid=1 Class"
          LvU="LvU: Low=0 vs Upp=1 Class";
  * Center age at 18 (minimum in sample);
    age18=age-18;
    LABEL age18= "age18: Age (0=18 years)";
  * 3 Piecewise slopes for education;
    lessHS=.; gradHS=.; overHS=.; * Make three new empty variables;
    * Replace for educ less than 12;
      IF educ LT 12 THEN DO; lessHS=educ-11; gradHS=0; overHS=0;       END;
    * Replace for educ greater or equal to 12;
      IF educ GE 12 THEN DO; lessHS=0;       gradHS=1; overHS=educ-12; END;
      LABEL lessHS= "lessHS: Slope for Years Ed Less Than High School"
            gradHS= "gradHS: Bump for Graduating High School"
            overHS= "overHS: Slope for Years Ed After High School";
* Label outcome;
  LABEL income= "income: Annual Income in 1000s";
* Select cases complete on variables;
  WHERE NMISS(income,educ,marry,workclass,age)=0;
RUN;
* Now dataset work.Example4a is ready to use;
```

STATA Syntax for Importing and Preparing Data for Analysis:

```
// Paste in the folder address where "GSS_Example.xlsx" is saved between " "
cd "\\Client\C:\Dropbox\22SP_PSQF6243\PSQF6243_Example4a"

// IMPORT GSS_Example.xlsx data from working directory using exact file name
// To change all variable names to lowercase, remove "case(preserve")
clear // Clear before means close any open data
import excel "GSS_Example.xlsx", case(preserve) firstrow clear
// Clear after means re-import if it already exists (if need to start over)

// Create and label predictor variables for model 1 (same as in Example 2)
// Linear education predictor centered so that 0 is meaningful
gen educ12=educ-12
label variable educ12 "educ12: Education (0=12 years)"
// Recode marry predictor so that 0 is meaningful
gen marry01=. // Create new empty variable, then recode
replace marry01=0 if marry==1
replace marry01=1 if marry==2
label variable marry01 "marry01: 0=unmarried, 1=married"

// Create and label predictor variables for model 2 (same as in Example 3)
// 2 Indicator-dummy-coded binary predictors for workclass
gen LvM=. // Make two new empty variables
gen LvU=.
replace LvM=0 if workclass==1 // Replace each for lower
replace LvU=0 if workclass==1
replace LvM=1 if workclass==2 // Replace each for middle
replace LvU=0 if workclass==2
replace LvM=0 if workclass==3 // Replace each for upper
replace LvU=1 if workclass==3
label variable LvM "LvM: Low=0 v Mid=1 Class"
label variable LvU "LvU: Low=0 v Upp=1 Class"
// Center age at 18 (minimum in sample)
gen age18 = age-18
label variable age18 "age18: Age (0=18 years)"
// 3 Piecewise slopes for education
gen lessHS=. // Make 3 new empty variables
gen gradHS=.
gen overHS=.
// Replace for educ less than 12
replace lessHS=educ-11 if educ <  12
replace gradHS=0 if educ <  12
replace overHS=0 if educ <  12
// Replace for educ greater or equal to 12
replace lessHS=0 if educ >= 12
replace gradHS=1 if educ >= 12
replace overHS=educ-12 if educ >= 12
// Label variables
label variable lessHS "lessHS: Slope for Years Ed Less Than High School"
label variable gradHS "gradHS: Acute Bump for Graduating High School"
label variable overHS "overHS: Slope for Years Ed After High School"
// Label outcome
label variable income "income: Annual Income in 1000s"
// Select cases complete on variables of interest
egen nmiss = rowmiss(income workclass age educ)
drop if nmiss>0
// Now dataset is ready to use
```

R Syntax for Importing and Preparing Data for Analysis:

```
# Set working directory (to import and export files to)
# Paste in the folder address where "GSS_Example.xlsx" is saved in quotes
setwd("C:/Dropbox/22SP_PSQF6243/PSQF6243_Example4a")

# Import GSS_Example.xlsx data from working directory -- path = file name
Example4a = read_excel(path="GSS_Example.xlsx", sheet="GSS_Example")
# Convert to data frame to use for analysis
Example4a = as.data.frame(Example4a)

### Create and label predictor variables for model 1 (same as in Example 2)
# Linear predictor for education centered so that 0 is meaningful
Example4a$educ12=Example4a$educ-12
# educ12: Education (0=12 years)
# Recode marry predictor so that 0 is meaningful
Example4a$marry01=NA  # Create new empty variable, then recode
Example4a$marry01[which(Example4a$marry==1)]=0
Example4a$marry01[which(Example4a$marry==2)]=1
# marry01: 0=unmarried, 1=married

### Create and label predictor variables for model 2 (same as in Example 3)
# 2 Indicator-dummy-coded binary predictors for workclass
Example4a$LvM=NA; Example4a$LvU=NA  # Make 2 new empty variables
Example4a$LvM[which(Example4a$workclass==1)]=0  # Replace each for lower
Example4a$LvU[which(Example4a$workclass==1)]=0
Example4a$LvM[which(Example4a$workclass==2)]=1  # Replace each for middle
Example4a$LvU[which(Example4a$workclass==2)]=0
Example4a$LvM[which(Example4a$workclass==3)]=0  # Replace each for upper
Example4a$LvU[which(Example4a$workclass==3)]=1
# LvM: Low=0 vs Mid=1 Class
# LvU: Low=0 vs Upp=1 Class
# Center age at 18 (minimum in sample)
Example4a$age18=Example4a$age-18
# age18: Age (0=18 years)
# Make squared age for GLHT statements
Example4a$agesq=Example4a$age18*Example4a$age18
# agesq: Squared Age (0=18 years)
# 3 Piecewise slopes for education
# Make 3 new empty variables
Example4a$lessHS=NA; Example4a$gradHS=NA; Example4a$overHS=NA
# Replace each for educ less than 12
Example4a$lessHS[which(Example4a$educ<12)]=Example4a$educ[which(Example4a$educ<12)]-11
Example4a$gradHS[which(Example4a$educ<12)]=0
Example4a$overHS[which(Example4a$educ<12)]=0
# Replace each for educ greater or equal to 12
Example4a$lessHS[which(Example4a$educ>=12)]=0
Example4a$gradHS[which(Example4a$educ>=12)]=1
Example4a$overHS[which(Example4a$educ>=12)]=Example4a$educ[which(Example4a$educ>=12)]-12
# lessHS: Slope for Years Ed Less Than High School
# gradHS: Acute Bump for Graduating High School
# overHS: Slope for Years Ed After High School
# Label outcome variable
# income: Annual Income in 1000s
# Now Example4a dataset is ready to use
```

Model 1: Linear Education and Binary Marital Status Predicting Income

Below is a summary of the results from estimating separate models per conceptual predictor (as demonstrated in example 2, as well as in the syntax and output online). Because there was only one conceptual predictor in each model, the model R^2 = semi-partial R^2 = partial R^2 (see excel sheet available online for calculations).

Separate Models	DF num	Effect SS	Residual SS	Total SS	semi-partial R2	partial R2
Empty Model	0	0	139423	139423	0.0000	0.0000
Linear Education	1	20635	118788	139423	0.1480	0.1480
Binary Marital Status	1	7060	132363	139423	0.0506	0.0506
Sum of Separate Models	2	27695			0.1986	0.1986

Combined: $Income_i = \beta_0 + \beta_1(educ_i - 12) + \beta_2(marry01_i) + e_i$

```
TITLE "SAS Model 1: Predict Income from Linear Education and Binary Marital Status";
PROC GLM DATA=work.Example4a NAMELEN=100;
     MODEL income = educ12 marry01 / SOLUTION ALPHA=.05 CLPARM SS3 EFFECTSIZE;
RUN; QUIT; TITLE;

display "STATA Model 1: Predict Income from Linear Education and Binary Marital Status"
regress income c.educ12 c.marry01, level(95)
display "STATA Semi-Partial and Partial Effect Sizes"
pcorr income educ12 marry01

print("R Model 1: Predict Income from Linear Education and Binary Marital Status")
Model1 = lm(data=Example4a, formula=income~1+educ12+marry01)
anova(Model1)   # anova to print residual variance
SumModel1 = summary(Model1); SumModel1 # save summary and print fixed effects solution
confint(Model1, level=.95) # confint for level% CI for fixed effects

print("R Partial and Semi-Partial R2 of income with educ12")
 pcor.test(Example4a$income, Example4a$educ12, Example4a[,"marry01"])$estimate^2
spcor.test(Example4a$income, Example4a$educ12, Example4a[,"marry01"])$estimate^2
print("R Partial and Semi-Partial R2 of income with marry01")
 pcor.test(Example4a$income, Example4a$marry01, Example4a[,"educ12"])$estimate^2
spcor.test(Example4a$income, Example4a$marry01, Example4a[,"educ12"])$estimate^2
```

Partial SAS Output:

```
SAS Model 1: Predict Income from Linear Education and Binary Marital Status
                                            Sum of
Source                      DF             Squares     Mean Square    F Value    Pr > F
Model                        2          26530.4118      13265.2059      85.89    <.0001
Error                      731         112892.8201        154.4361
Corrected Total            733         139423.2319

R-Square     Coeff Var      Root MSE    income Mean
0.190287      71.82179      12.42723       17.30287
```

Mean Square Error, the residual variance, is 154.44 after including 1 slope each for the 2 predictor constructs (which accounted for 19.03% of the variance in income as the model R^2). The *F*-test says this R^2 is significantly > 0, $F(2, 731) = 85.89$, MSE = 154.44, $p < .001$.

New effect size output

```
                Total Variation Accounted For    Partial Variation Accounted For
                                 Semipartial                            Partial
                 Semipartial          Omega-           Partial           Omega-
Source            Eta-Square          Square        Eta-Square           Square
educ12                0.1396          0.1384            0.1471           0.1456
marry01               0.0423          0.0411            0.0496           0.0482
```

```
                                        Standard
Parameter          Estimate             Error     t Value    Pr > |t|      95% Confidence Limits
Intercept       11.47412468        0.67752192       16.94      <.0001     10.14400382  12.80424555  Beta0
educ12           1.77384962        0.15798099       11.23      <.0001      1.46369905   2.08400020  Beta1
marry01          5.69460677        0.92168064        6.18      <.0001      3.88514996   7.50406358  Beta2
```

Partial STATA Output:

```
      Source |       SS           df       MS      Number of obs   =       734
-------------+----------------------------------   F(2, 731)       =     85.89
       Model |  26530.4118         2  13265.2059   Prob > F        =    0.0000
    Residual |   112892.82       731  154.436142   R-squared       =    0.1903
-------------+----------------------------------   Adj R-squared   =    0.1881
       Total |  139423.232       733  190.209048   Root MSE        =    12.427

------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      educ12 |    1.77385    .157981    11.23   0.000     1.463699       2.084
     marry01 |   5.694607   .9216806     6.18   0.000      3.88515    7.504064
       _cons |   11.47412   .6775219    16.94   0.000       10.144    12.80425
------------------------------------------------------------------------------

Partial and semipartial correlations of income with
               Partial   Semipartial      Partial   Semipartial   Significance
   Variable |    Corr.         Corr.      Corr.^2       Corr.^2          Value
------------+-----------------------------------------------------------------
     educ12 |   0.3835        0.3737       0.1471        0.1396         0.0000
    marry01 |   0.2228        0.2056       0.0496        0.0423         0.0000
```

Partial R Output:

```
Analysis of Variance Table
Response: income
           Df   Sum Sq  Mean Sq  F value          Pr(>F)
educ12      1  20635.0 20634.98 133.6150      < 2.22e-16
marry01     1   5895.4  5895.43  38.1739 0.0000000010743
Residuals 731 112892.8   154.44
> SumModel1 = summary(Model1)
> SumModel1
```

Mean Sq (Square) for "Residuals"
= Residual Variance = 154.44

```
Coefficients:
            Estimate Std. Error t value       Pr(>|t|)
(Intercept) 11.47412    0.67752 16.9354      < 2.2e-16  Beta0
educ12       1.77385    0.15798 11.2282      < 2.2e-16  Beta1
marry01      5.69461    0.92168  6.1785 0.000000001074  Beta2

Residual standard error: 12.427 on 731 degrees of freedom
Multiple R-squared:  0.19029, Adjusted R-squared:  0.18807
F-statistic: 85.894 on 2 and 731 DF,  p-value: < 2.22e-16

[1] "R Partial and Semi-Partial R2 of income with educ12"
[1] 0.14709769
[1] 0.13964897

[1] "R Partial and Semi-Partial R2 of income with marry01"
[1] 0.049629741
[1] 0.042284417
```

Below is a summary of the results for the overall model and the contribution of each predictor— As shown in the last row above, the sum across the two constructs of the Effects Sums of Squares (SS) differs from the Model SS—this is because the Model SS takes into account the predictive contribution of the shared variance among the sets of predictors (i.e., none of them "gets credit" for what they have in common that predicts income, but the model R^2 does reflect that common contribution).

Combined Model	DF num	Effect SS	Residual SS	Total SS	semi-partial R2	partial R2
Full Model	2	26530	112893	139423	0.1903	0.1903
Linear Education	1	19470	112893	139423	0.1396	0.1471
Binary Marital Status	1	5895	112893	139423	0.0423	0.0496
Sum of Predictors		25366			0.1819	0.1967

See syntax and output online for how to get partial d, partial r, and semi-partial r using t-values:

Effect from Construct-Separate Models	Est	SE	t	p	DF den	Partial d	Partial r
Linear Education	1.824	0.162	11.276	<.0001	732	0.8336	0.3847
Binary Marital Status	6.224	0.996	6.249	<.0001	732	0.4619	0.2250
Effect from Combined Model	**Est**	**SE**	**t**	**p**	**DF den**	**Partial d**	**Partial r**
Linear Education	1.774	0.158	11.228	<.0001	731	0.8306	0.3835
Binary Marital Status	5.695	0.922	6.179	<.0001	731	0.4570	0.2228

Model 2: Three-Category Workclass, Linear and Quadratic Age Slopes, and Three Piecewise Linear Slopes for Education Predicting Income

Below is a summary of the results from estimating separate models per conceptual predictor (as demonstrated in example 3, as well as in the syntax and output online). Because there was only one conceptual predictor in each model, the model R^2 = semi-partial R^2 = partial R^2 (see excel sheet available online for calculations).

Separate Models	DF num	Effect SS	Residual SS	Total SS	semi-partial R2	partial R2
3-Category Workclass	2	14414	125009	139423	0.1034	0.1034
Linear + Quadratic Age	2	15885	123538	139423	0.1139	0.1139
3 Piecewise Slopes Education	3	22907	116517	139423	0.1643	0.1643
Sum of Separate Models	2	53206			0.3816	0.3816

Combined: $$Income_i = \beta_0 + \beta_1(LvsM_i) + \beta_2(LvsU_i) + \beta_3(Age_i - 18) + \beta_4(Age_i - 18)^2 + \beta_5(LessHS_i) + \beta_6(GradHS_i) + \beta_7(OverHS_i) + e_i$$

In addition to the overall F-test of the model R^2, the purpose of estimating a single model with the seven slopes from all three predictive constructs combined (workclass, age, and education) is to determine to what extent their bivariate effects (when each construct was in a separate model predicting income, as was the case in Example 3) differ from their unique effects (when all constructs are combined in the same model, below). The solution for the fixed effects will provide tests for the significance of each slope (against a null hypothesis of a 0 slope in the population), and we will also ask for joint F-tests (and their effect sizes) that combine the multiple slopes needed to capture the full effect of each construct. SAS CONTRAST will provide effect sizes for each conceptual predictor along with the F-test, but effect sizes must be computed manually in STATA and R.

```
TITLE "SAS Model 2: Workclass, Quadratic Age, and Piecewise Education";
PROC GLM DATA=work.Example4a NAMELEN=100 PLOTS(UNPACK)=DIAGNOSTICS;
* Combined model with all 7 slopes;
  MODEL income = LvM LvU age18 age18*age18 lessHS gradHS overHS
                / SOLUTION ALPHA=.05 CLPARM SS3 EFFECTSIZE;
* Ask for missing model-implied group difference;
  ESTIMATE "Mid v Upp Diff"  LvM -1 LvU 1;

* Replicate F-test and R2 for the model: includes all 7 slopes;
  CONTRAST "F-test (DFnum=7) for model"
            LvM 1, LvU 1, age18 1, age18*age18 1, lessHS 1, gradHS 1, overHS 1;
* Ask for F-test and semi-partial R2 for overall effect of workclass;
  CONTRAST "F-test (DFnum=2) for overall workclass" LvM 1, LvU 1;
* Ask for F-test and semi-partial R2 for overall effect of age;
  CONTRAST "F-test (DFnum=2) for overall age" age18 1, age18*age18 1;
* Ask for F-test and semi-partial R2 for overall effect of education;
  CONTRAST "F-test (DFnum=3) for overall education" lessHS 1, gradHS 1, overHS 1;
RUN; QUIT; TITLE;
```

```
display "STATA Model 2: Workclass, Quadratic Age, and Piecewise Education"
regress income c.LvM c.LvU c.age18 c.age18#c.age18 c.lessHS c.gradHS c.overHS, level(95)
// Ask for missing model-implied group difference
lincom c.LvM*-1 + c.LvU*1 // Mid v Upp Diff

// Replicate F-test for the model: includes all 7 slopes
test (c.LvM=0)(c.LvU=0)(c.age18=0)(c.age18#c.age18=0)(c.lessHS=0)(c.gradHS=0)(c.overHS=0)
// Ask for F-test for overall effect of workclass
test (c.LvM=0)(c.LvU=0)
// Ask for F-test for overall effect of age
test (c.age18=0)(c.age18#c.age18=0)
// Ask for F-test for overall effect of education
test (c.lessHS=0)(c.gradHS=0)(c.overHS=0)
```

```
print("R Model 2: Predict Income from Workclass, Quadratic Age, and Piecewise Education")
Model2 = lm(data=Example4a, formula=income~1+LvM+LvU+age18+agesq+lessHS+gradHS+overHS)
anova(Model2)   # anova to print residual variance
SumModel2 = summary(Model2); SumModel2 # save summary and print fixed effects solution
confint(Model2, level=.95) # confint for level% CI for fixed effects

print("R Ask for missing model-implied group difference")
PredModel2 = glht(model=Model2, linfct=rbind("Mid vs Upp Diff" = c(0,-1,1,0,0,0,0,0)))
print("Save glht linear combination results with unadjusted p-values and 95% CIs")
SaveModel2 = summary(PredModel2, test=adjusted("none"))
print(SaveModel2); confint(PredModel2, level=.95, calpha=univariate_calpha())

print("Replicate F-test for model: includes all 7 slopes")
Model2Fall = glht(model=Model2, linfct=c("LvM=0","LvU=0","age18=0","agesq=0",
                                         "lessHS=0","gradHS=0","overHS=0"))
summary(Model2Fall, test=Ftest()) # ask for joint hypothesis test instead of separate

print("Ask for F-test for overall effect of workclass")
Model2Fclass = glht(model=Model2, linfct=c("LvM=0","LvU=0"))
summary(Model2Fclass, test=Ftest()) # ask for joint hypothesis test instead of separate

print("Ask for F-test for overall effect of age")
Model2Fage = glht(model=Model2, linfct=c("age18=0","agesq=0"))
summary(Model2Fage, test=Ftest()) # ask for joint hypothesis test instead of separate

print("Ask for F-test for overall effect of education")
Model2Feduc = glht(model=Model2, linfct=c("lessHS=0","gradHS=0","overHS=0"))
summary(Model2Feduc, test=Ftest()) # ask for joint hypothesis test instead of separate
```

Partial SAS Output:

SAS Model 2: Workclass, Quadratic Age, and Piecewise Education

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	**7**	40246.4243	5749.4892	**42.09**	<.0001
Error	**726**	99176.8076	**136.6072**		
Corrected Total	733	139423.2319			

R-Square	Coeff Var	Root MSE	income Mean
0.288664	67.54893	11.68791	17.30287

Mean Square Error, the residual variance, is 136.61 after including 7 slopes for the 3 predictor constructs (which accounted for 28.87% of the variance in income as the model R^2). The *F*-test says this R^2 is significantly > 0, $F(7, 726) = 42.04$, MSE = 136.61, $p < .001$.

	Total Variation Accounted For		Partial Variation Accounted For		
Source	**Semipartial Eta-Square**	Semipartial Omega-Square	Partial Eta-Square	Partial Omega-Square	
LvM	0.0401	0.0391	0.0534	0.0516	→ do not use
LvU	0.0070	0.0060	0.0097	0.0083	→ do not use
age18	0.0741	0.0731	0.0944	0.0923	→ ok, but conditional on 18
age18*age18	0.0580	0.0570	0.0754	0.0735	→ ok by itself
lessHS	0.0002	-0.0008	0.0003	-0.0011	→ ok by itself
gradHS	0.0032	0.0022	0.0044	0.0030	→ ok by itself
overHS	0.0529	0.0518	0.0692	0.0673	→ ok by itself

Because the workclass predictors are related (each shares a reference group with another), the total of the sr^2 values for these three differences they imply (two of which are given here) is greater than it should be, so these effect sizes are not valid. The linear age slope's effect size is valid but conditional on age 18. The per-slope effect sizes for the remainder of the slopes are ok, but these cannot be added together directly to represent the contribution per conceptual predictor. Instead, we need to obtain an *F*-test and effect size that combines the slopes for the same construct… that's what the CONTRAST statements were for! Let's see what they give us…

Tables from CONTRAST statements

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
F-test (DFnum=7) for model	7	40246.42433	5749.48919	42.09	<.0001
F-test (DFnum=2) for overall workclass	2	5961.75558	2980.87779	21.82	<.0001
F-test (DFnum=2) for overall age	2	11223.60775	5611.80388	41.08	<.0001
F-test (DFnum=3) for overall education	3	11251.78823	3750.59608	27.46	<.0001

	Total Variation Accounted For		Partial Variation Accounted For	
Contrast	**Semipartial Eta-Square**	Semipartial Omega-Square	Partial Eta-Square	Partial Omega-Square
F-test (DFnum=7) for model	0.2887	0.2815	0.2887	0.2815
F-test (DFnum=2) for overall workclass	0.0428	0.0408	0.0567	0.0537
F-test (DFnum=2) for overall age	0.0805	0.0785	0.1017	0.0985
F-test (DFnum=3) for overall education	0.0807	0.0777	0.1019	0.0976

The sr^2 values above give the amount of variance accounted for <u>each set of slopes</u> (the sets we requested using CONTRAST statements). Whether those sr^2 values are > 0 is tested by their corresponding *F*-value above.

Table from ESTIMATE statement (for model-implied fixed effects—is beta2 − beta1 here)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Mid v Upp Diff	**1.14843248**	2.70813034	0.42	0.6716	-4.16826904	6.46513400

Table of Model-Estimated Fixed Effects (normally is last)

Parameter	Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits		
Intercept	-3.686546177	2.00461546	-1.84	0.0663	-7.622081294	0.248988941	**Beta0**
LvM	6.060105402	0.94700667	6.40	<.0001	4.200906929	7.919303874	**Beta1**
LvU	7.208537879	2.69787938	2.67	0.0077	1.911961423	12.505114336	**Beta2**
age18	1.069979988	0.12300458	8.70	<.0001	0.828492845	1.311467130	**Beta3**
age18*age18	-0.017506167	0.00227492	-7.70	<.0001	-0.021972365	-0.013039969	**Beta4**
lessHS	0.258917912	0.56120164	0.46	0.6447	-0.842853869	1.360689693	**Beta5**
gradHS	3.157139208	1.75726664	1.80	0.0728	-0.292791564	6.607069980	**Beta6**
overHS	1.528179214	0.20804233	7.35	<.0001	1.119742828	1.936615600	**Beta7**

Partial STATA Output:

STATA GLM Predicting Income from Workclass, Age, and Education

```
      Source |       SS           df       MS      Number of obs   =       734
-------------+----------------------------------   F(7, 726)       =     42.09
       Model |  40246.4243         7  5749.48919   Prob > F        =    0.0000
    Residual |  99176.8076       726  136.607173   R-squared       =    0.2887
-------------+----------------------------------   Adj R-squared   =    0.2818
       Total |  139423.232       733  190.209048   Root MSE        =    11.688
--------------------------------------------------------------------------------
         income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
            LvM |   6.060105   .9470067     6.40   0.000     4.200907    7.919304   Beta1
            LvU |   7.208538   2.697879     2.67   0.008     1.911961    12.50511   Beta2
          age18 |    1.06998   .1230046     8.70   0.000     .8284928    1.311467   Beta3
c.age18#c.age18 |  -.0175062   .0022749    -7.70   0.000    -.0219724    -.01304   Beta4
         lessHS |   .2589179   .5612016     0.46   0.645    -.8428539     1.36069   Beta5
         gradHS |   3.157139   1.757267     1.80   0.073    -.2927916     6.60707   Beta6
         overHS |   1.528179   .2080423     7.35   0.000     1.119743    1.936616   Beta7
          _cons |  -3.686546   2.004615    -1.84   0.066    -7.622081    .2489889   Beta0
--------------------------------------------------------------------------------
. // Ask for model-implied group difference
.    lincom  c.LvM*-1 + c.LvU*1 // Mid v Upp Diff
------------------------------------------------------------------------------
      income |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   1.148432    2.70813     0.42   0.672    -4.168269    6.465134   Beta2 - Beta1
------------------------------------------------------------------------------
```

From new STATA TEST statements for custom F-tests:

```
. // Replicate F-test for the model: includes all 7 slopes
 ( 1)  LvM = 0
 ( 2)  LvU = 0
 ( 3)  age18 = 0
 ( 4)  c.age18#c.age18 = 0
 ( 5)  lessHS = 0
 ( 6)  gradHS = 0
 ( 7)  overHS = 0
       F(  7,   726) =   42.09
            Prob > F =    0.0000

. // Ask for F-test for overall effect of workclass
 ( 1)  LvM = 0
 ( 2)  LvU = 0
       F(  2,   726) =   21.82
            Prob > F =    0.0000

. // Ask for F-test for overall effect of age
 ( 1)  age18 = 0
 ( 2)  c.age18#c.age18 = 0
       F(  2,   726) =   41.08
            Prob > F =    0.0000

. // Ask for F-test for overall effect of education
 ( 1)  lessHS = 0
 ( 2)  gradHS = 0
 ( 3)  overHS = 0
```

```
       F(  3,   726) =   27.46
            Prob > F =    0.0000
```

Partial R Output:

```
Analysis of Variance Table
Response: income
           Df  Sum Sq  Mean Sq F value             Pr(>F)
LvM         1 12106.6 12106.59 88.6234          < 2.22e-16
LvU         1  2307.4  2307.43 16.8910 0.00004410462407414
age18       1  3685.4  3685.40 26.9781 0.00000026756426081
agesq       1 10895.2 10895.21 79.7558          < 2.22e-16
lessHS      1  1844.0  1843.97 13.4983          0.00025634
gradHS      1  2037.0  2036.95 14.9110          0.00012276
overHS      1  7370.9  7370.87 53.9567 0.00000000000055215
Residuals 726 99176.8   136.61
```

Mean Sq (Square) for "Residuals" = Residual Variance = 136.61

```
Coefficients:
              Estimate Std. Error t value  Pr(>|t|)
(Intercept) -3.6865462  2.0046155 -1.8390  0.066319  Beta0
LvM          6.0601054  0.9470067  6.3992 2.800e-10  Beta1
LvU          7.2085379  2.6978794  2.6719  0.007711  Beta2
age18        1.0699800  0.1230046  8.6987 < 2.2e-16  Beta3
agesq       -0.0175062  0.0022749 -7.6953 4.615e-14  Beta4
lessHS       0.2589179  0.5612016  0.4614  0.644676  Beta5
gradHS       3.1571392  1.7572666  1.7966  0.072812  Beta6
overHS       1.5281792  0.2080423  7.3455 5.522e-13  Beta7

Residual standard error: 11.688 on 726 degrees of freedom
Multiple R-squared:  0.28866,        Adjusted R-squared:  0.28181
F-statistic: 42.088 on 7 and 726 DF,  p-value: < 2.22e-16
```

From new R GLHT statements (for linear combinations or custom F-tests):

```
Linear Hypotheses:
                     Estimate lwr      upr
Mid vs Upp Diff == 0  1.14843 -4.16827  6.46513  Beta2 - Beta1

[1] "Replicate F-test for model: includes all 7 slopes
Linear Hypotheses:
             Estimate
LvM == 0     6.060105
LvU == 0     7.208538
age18 == 0   1.069980
agesq == 0  -0.017506
lessHS == 0  0.258918
gradHS == 0  3.157139
overHS == 0  1.528179
Global Test:
       F DF1 DF2     Pr(>F)
1 42.088   7 726 6.9979e-50

[1] "Ask for F-test for overall effect of workclass"
Linear Hypotheses:
         Estimate
LvM == 0   6.0601
LvU == 0   7.2085
Global Test:
       F DF1 DF2           Pr(>F)
1 21.821   2 726 0.00000000062698

[1] "Ask for F-test for overall effect of age"
Linear Hypotheses:
            Estimate
age18 == 0  1.069980
agesq == 0 -0.017506
Global Test:
      F DF1 DF2     Pr(>F)
1 41.08   2 726 1.2546e-17

[1] "Ask for F-test for overall effect of education"
Linear Hypotheses:
```

```
                 Estimate
lessHS == 0   0.25892
gradHS == 0   3.15714
overHS == 0   1.52818
Global Test:
       F DF1 DF2    Pr(>F)
1 27.455   3 726 7.95e-17
```

Below is a summary of the results for the overall model and the contribution of each predictor*:*

Combined Model	DF num	Effect SS	Residual SS	Total SS	semi-partial R2	partial R2
Full Model	7	40246	99177	139423	0.2887	0.2887
2 Group Differences Workclass	2	5962	99177	139423	0.0428	0.0567
Linear + Quadratic Age	2	11224	99177	139423	0.0805	0.1017
3 Piecewise Slopes Education	3	11252	99177	139423	0.0807	0.1019
Sum of Workclass, Age, Education	7	28437			0.2040	0.2603

As shown in the last row above, the sum across the three constructs of the Effects Sums of Squares (SS) differs from the Model SS—this is because the Model SS takes into account the predictive contribution of the shared variance among the sets of predictors (i.e., none of them "gets credit" for what they have in common that predicts income, but the model R^2 does reflect that common contribution).

Here is how to get the effect sizes (as given directly by SAS CONTRAST) manually in STATA and R by finding the sums of squares for the unique contribution of each conceptual predictor:

```
display "STATA Reduced Model to Get SS for workclass (not included)"
regress income c.age18 c.age18#c.age18 c.lessHS c.gradHS c.overHS, level(95)

      Source |       SS           df       MS      Number of obs   =       734
-------------+----------------------------------   F(5, 728)       =     47.48
       Model |  34284.6688         5  6856.93375   Prob > F        =    0.0000
    Residual |  105138.563       728  144.421103   R-squared       =    0.2459
-------------+----------------------------------   Adj R-squared   =    0.2407
       Total |  139423.232       733  190.209048   Root MSE        =    12.018

// sr2 = (SSfull-SSreduced)/SStotal
display (40246.42433-34284.6688)/139423.2319 // sr2 for workclass
.04276013
```

The R version uses SStotal saved from the empty model first:

```
print("R Empty Model Predicting Income")
ModelEmpty = lm(data=Example4b, formula=income~1)
anova(ModelEmpty)    # anova to print residual variance
summary(ModelEmpty) # summary to print fixed effects solution
confint(ModelEmpty, level=.95) # confint for level% CI for fixed effects

# Save sums of squares from empty model for later calculations
SStotal=anova(ModelEmpty)$`Sum Sq`


print("R Semi-Partial R2 for Workclass via SS for Reduced Model Omitting Workclass")
Model2NoClass = lm(data=Example4a, formula=income~1+age18+agesq+lessHS+gradHS+overHS)
SSClass=anova(Model2NoClass, Model2); SSClass$`Sum of Sq`/SStotal
NA 0.04276013
```

```
display "STATA Reduced Model to Get SS for age (not included)"
regress income c.LvM c.LvU c.lessHS c.gradHS c.overHS, level(95)

      Source |       SS           df       MS      Number of obs   =       734
-------------+----------------------------------   F(5, 728)       =     38.28
       Model |  29022.8166         5  5804.56332   Prob > F        =    0.0000
    Residual |  110400.415       728  151.648922   R-squared       =    0.2082
-------------+----------------------------------   Adj R-squared   =    0.2027
       Total |  139423.232       733  190.209048   Root MSE        =    12.315

// sr2 = (SSfull-SSreduced)/SStotal
display (40246.42433-29022.8166)/139423.2319 // sr2 for age
.08050027
```

```
print("R Semi-Partial R2 for Age via SS for Reduced Model Omitting Age")
Model2NoAge = lm(data=Example4a, formula=income~1+LvM+LvU+lessHS+gradHS+overHS)
SSAge=anova(Model2NoAge, Model2); SSAge$`Sum of Sq`/SStotal
NA 0.08050027
```

```
display "STATA Reduced Model to Get SS for education (not included)"
regress income c.LvM c.LvU c.age18 c.age18#c.age18, level(95)

      Source |       SS           df       MS      Number of obs   =       734
-------------+----------------------------------   F(4, 729)       =     47.85
       Model |  28994.6361         4  7248.65903   Prob > F        =    0.0000
    Residual |  110428.596       729  151.479555   R-squared       =    0.2080
-------------+----------------------------------   Adj R-squared   =    0.2036
       Total |  139423.232       733  190.209048   Root MSE        =    12.308

// sr2 = (SSfull-SSreduced)/SStotal
display (40246.42433-28994.6361)/139423.2319 // sr2 for education
.08070239
```

```
print("R Semi-Partial R2 for Education via SS for Reduced Model Omitting Education")
Model2NoEduc = lm(data=Example4a, formula=income~1+LvM+LvU+age18+agesq)
SSEduc=anova(Model2NoEduc, Model2); SSEduc$`Sum of Sq`/SStotal
NA 0.080702391
```

Here is a comparison of the results from each construct in a separate model (from Example 3) with the present results from a combined model with all 7 slopes—*see syntax and output online for how to get partial d, partial r, and semi-partial r using t-values*

Effect from Construct-Separate Models	Est	SE	t	p	DF den	d	r
Lower vs Middle Class	8.854	1.004	8.822	<.0001	731	0.65	0.310
Lower vs Upper Class	10.985	2.990	3.673	0.000	731	0.27	0.135
(Middle vs Upper Class)	2.130	3.027	0.704	0.482	731	0.05	0.026
Linear Age Slope	1.223	0.135	9.055	<.0001	731	0.67	0.318
Quadratic Age Slope	-0.020	0.003	-7.809	<.0001	731	-0.58	-0.277
Education 2 to 11 years	-0.269	0.599	-0.449	0.654	730	-0.03	-0.017
Education: 11 to 12 years	4.685	1.876	2.498	0.013	730	0.18	0.092
Education: 12 to 20 years	2.125	0.214	9.941	<.0001	730	0.74	0.345
Effect from Combined Model	**Est**	**SE**	**t**	**p**	**DF den**	**d**	**r**
Lower vs Middle Class	6.060	0.947	6.400	<.0001	726	0.47	0.231
Lower vs Upper Class	7.209	2.698	2.670	0.008	726	0.20	0.099
(Middle vs Upper Class)	1.148	2.708	0.420	0.672	726	0.03	0.016
Linear Age Slope	1.070	0.123	8.700	<.0001	726	0.65	0.307
Quadratic Age Slope	-0.018	0.002	-7.700	<.0001	726	-0.57	-0.275
Education 2 to 11 years	0.259	0.561	0.460	0.645	726	0.03	0.017
Education: 11 to 12 years	3.157	1.757	1.800	0.073	726	0.13	0.067
Education: 12 to 20 years	1.528	0.208	7.350	<.0001	726	0.55	0.263
Difference: Separate Minus Combined	**Est**					**d**	**r**
Lower vs Middle Class	2.794					0.178	0.079
Lower vs Upper Class	3.776					0.073	0.036
(Middle vs Upper Class)	0.982					0.021	0.010
Linear Age Slope	0.153					0.024	0.010
Quadratic Age Slope	-0.002					-0.006	-0.003
Education 2 to 11 years	-0.528					-0.067	-0.034
Education: 11 to 12 years	1.528					0.052	0.026
Education: 12 to 20 years	0.596					0.191	0.082

Example Results Section for Model 2 (would continue from separate results described in Example 3):

[Table 1 would report the parameter estimates from the combined model, along with partial *d* and *r* effect sizes. The table note would indicate how they were computed: $d = \frac{2t}{\sqrt{DF_{den}}}$; $r = \frac{t}{\sqrt{t^2+DF_{den}}}$]

After examining the bivariate contributions of three-category self-reported working class membership, linear and quadratic years of age, and piecewise slopes for years of education in separate models, we then estimated a combined model to examine their unique contributions after controlling for each other predictor variable. The model including all seven slopes captured a significant amount of variance in annual income, $F(7, 726) = 42.09$, MSE = 136.61, $p < .001$, $R^2 = .289$. Parameter estimates and effect sizes are given in Table 1. Semipartial eta-squared (η^2) effect sizes and corresponding multivariate Wald *F*-tests were obtained to evaluate the amount of total variance captured by distinct sets of predictor slopes.

The omnibus unique effect of three-category self-reported working class membership remained significant, $F(2, 726) = 21.83$, MSE = 136.61, $p < .0001$, semipartial $\eta^2 = .043$. As shown in Table 1, relative to lower-class respondents (the reference group), after controlling for years of age and years of education, annual income was still significantly higher for both middle-class and upper-class respondents (by 6.060 and 7.209 thousand dollars, respectively). Middle-class and upper-class respondents still did not differ significantly in predicted annual income.

The omnibus unique effect of quadratic years of age (centered at 18) also remained significant, $F(2, 726) = 41.08$, MSE = 136.61, $p < .0001$, semipartial $\eta^2 = .081$. As shown in Table 1, after controlling for self-reported working class and years of education, annual income was expected to be significantly higher by 1.070 thousand dollars per year of age at age 18; this instantaneous linear age slope was predicted to become significantly less positive per year of age by twice the quadratic coefficient of −0.018. As given by the quantity (−1*linear slope) / (2*quadratic slope) + 18, the age of maximum predicted personal income was 48.56 (i.e., the age at which the linear age slope = 0).

The omnibus unique effect of piecewise years of education (centered at 11) also remained significant, $F(3, 726) = 27.46$, MSE = 136.61, $p < .0001$, semipartial $\eta^2 = .081$. As shown in Table 1, after controlling for self-reported working class and years of age, annual income was expected to be nonsignificantly higher by 0.259 thousand dollars per year of education from 2 to 11 years, to be nonsignificantly higher by 3.157 thousand dollars for those achieving a high school degree, and to be significantly higher by 1.528 thousand dollars per year of additional education past 12 years. Notably, the effect of a high school degree (the difference between 11 and 12 years of education) was no longer significant after controlling for age and self-reported working class membership.

[The rest of the text would need to emphasize why it matters based on your research questions that the predictors had significant unique effects. This is the part that must be customized per research study!]