Lecture Note 9

Modeling with Multivariate Regression Models

1 Saturated Regression Models and Interaction Terms

- LN5 (Sec.4) introduces the idea of *saturated regression models* – let's expand on this
  - A saturated regression model has as many parameters as the corresponding CEF
  - When counting parameters, we include the intercept
- Consider a single multinomial regressor $X_i \in \{10, 11, 12, ...19\}$. The CEF $E[Y_i|X_i]$ assumes up to 10 distinct values.
  - What's the corresponding saturated regression model look like? An intercept plus 9 dummies, one for each value of $X_i > 10$
  - Other 10-parameter schemes also work, provided the regression implementation is full-rank (has no linear dependencies)
- What about multiple regressors? Suppose X_{1i} and X_{2i} are dummies for college graduation status and sex (not that these ever go together):

$$\begin{aligned} X_{1i} &= 1(colgrad_i = yes) \\ X_{2i} &= 1(sex_i = female) \end{aligned}$$

- The CEF of Y_i conditional on X_{1i}, X_{2i} can be written:

$$E[Y_i \mid X_{1i}, X_{2i}] = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{12} X_{1i} X_{2i}$$

  How many values can this CEF assume?
  - The corresponding saturated regression model is

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{12} X_{1i} X_{2i} + \epsilon_i \tag{1}$$

- Think of saturated models as maximum flexibility
  - Equation (1) allows the college wage effect to differ for men and women:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \epsilon_i && \text{when } X_{2i} = 0 \\ Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 + \beta_{12} X_{1i} + \epsilon_i \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) X_{1i} + \epsilon_i && \text{when } X_{2i} = 1 \end{aligned}$$

  while also allowing the female effect to differ between those with and without a college degree::

$$\begin{aligned} Y_i &= \beta_0 + \beta_2 X_{2i} + \epsilon && \text{when } X_{1i} = 0 \\ Y_i &= \beta_0 + \beta_1 + \beta_2 X_{2i} + \beta_{12} X_{2i} + \epsilon_i \\ &= (\beta_0 + \beta_1) + (\beta_2 + \beta_{12}) X_{2i} + \epsilon && \text{when } X_{1i} = 1 \end{aligned}$$

  - What more could you ask for? (Nothing in this case)

- Regression talk
  - The coefficients on X_{1i} and X_{2i} are said to be *main effects*, while β_{12} is an *interaction term* or *2nd-order term.* Models with 3 or more regressors may have *higher-order terms*
  - Models without interactions are said to be *additive.* The additive version of (1) is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i \tag{2}$$

    Additive models can be viewed as either a restriction on, or an approximation to, the CEF
  - When choosing the details of a regression model by including or omitting interaction terms, we are said to be *specifying* or *parameterizing* it.
  - Critics of your model may object to your *specification* of it

2 Not Just for Dummies: Ordinal and Continuous Interactions

Interesting interactions arise in models mixing ordinal and continuous regressors.

- Let

$$S_i = years\ of\ schooling$$
$$X_{2i} = 1(sex_i = female)$$

- Repeat equation (1):

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_{2i} + \beta_{12} S_i X_{2i} + \epsilon_i \tag{3}$$

  - This model isn't saturated even though it includes an interaction term (why not?)
  - Regression model (3) therefore restricts or approximates the CEF (how?)

- With ordinal S_i and dummy X_{2i}, models like (3) are usually interpreted (asymmetrically) as allowing the regression of Y_i on S_i to differ by X_{2i}:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i && \text{when } X_{2i} = 0 \\ Y_i &= \beta_0 + \beta_1 S_i + \beta_2 + \beta_{12} S_i + \epsilon_i && \\ &= (\beta_0 + \beta_2) + (\beta_1 + \beta_{12}) S_i + \epsilon_i && \text{when } X_{i2} = 1 \end{aligned}$$

  - The (female-male) difference in intercepts is β_2
  - The (female-male) difference in the economic returns to schooling is β_{12}

- This parameterization produces results identical to those generated by separate regressions on S_i for men and women
  - Why? Both the S_i slope and the model intercept are free to vary with X_{2i}, which takes on two values
  - We say that model (3) is *fully interacted with* X_{2i} (although it's not saturated)

- A test of whether the regression on schooling is the same for men and women is a joint test of:

$$H_0 : \beta_2 = \beta_{12} = 0 \tag{4}$$

  - Multiple restrictions of this sort are tested with an F statistic (how many restrictions in (4)?).
  - You might be especially interested in testing $\beta_{12} = 0$, since, almost certainly, $\beta_2 \neq 0$

3 Testing Linear Restrictions

- We're often interested in testing sets of linear restrictions. A set of q linearly independent restrictions applied to a regression model with k coefficients $(\beta_1, ..., \beta_k)$ and an intercept (β_0) can be written:

$$c_{11}\beta_1 + c_{12}\beta_2 + c_{13}\beta_3 + ... + c_{1k}\beta_k = c_{10}$$
$$\vdots$$
$$c_{q1}\beta_1 + c_{q2}\beta_2 + c_{q3}\beta_3 + ... + c_{qk}\beta_k = c_{qo},$$

  where the c's are constants (we rarely restrict the intercept)

  - Rewriting (3) to match this notation:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_{2i} + \beta_3 S_i X_{2i} + \epsilon_i \quad (5)$$

    The restrictions in (4) can be written as imposing: $c_{12} = c_{23} = 1; c_{10} = c_{20} = 0$.

- Let e_{Ri} be the residuals from a model that imposes these restrictions and let e_{Ui} be the residuals from the model that doesn't impose them. Estimated residual variances from these models are $s_R^2 = \frac{1}{n-k-1}\sum e_{Ri}^2$ and $s_U^2 = \frac{1}{n-k-1}\sum e_{Ui}^2$. Given classical regression modeling assumptions (linear CEF, fixed regressors, homoskedastic Normal independent resids), we have that:

$$F = \frac{(\sum e_{Ri}^2 - \sum e_{Ui}^2)/q}{\sum e_{Ui}^2/(n-k-1)} = (\frac{n-k-1}{q}) \times \frac{(s_R^2 - s_U^2)}{s_U^2} \sim F_{q,\, n-k-1},$$

  under the null hypothesis that the restrictions are satisfied

  - This F *statistic* is used to do an F *test*
  - An F distribution has two parameters: *numerator degrees of freedom (df)* equal to the number of restrictions being tested and *denominator df* equal to the sample size minus the number of parameters in the unrestricted model
  - As usual, we test a null by looking for surprising draws from the null distribution, fixing the probability of Type I error
  - An F-stat's df determine the relevant critical values (numerator *df* matter most; denominator *df* can often be taken to be infinite)

- We are said to be testing *zero restrictions* where $c_{10=...} = c_{q0} = 0$ and all other c's are equal to either 1 or 0.

R^2 version of the F-test

- Write $R_R^2 = 1 - (s_R^2/s_Y^2)$ and $R_U^2 = 1 - (s_U^2/s_Y^2)$ for restricted and unrestricted R^2, respectively. Provided restrictions don't transform the dependent variable (thereby changing its variance), we can write:

$$F = (\frac{n-k-1}{q}) \times \frac{(s_R^2 - s_U^2)}{s_U^2} = (\frac{n-k-1}{q}) \times \frac{(s_R^2/s_Y^2 - s_U^2/s_Y^2)}{s_U^2/s_Y^2}$$
$$= \left(\frac{n-k-1}{q}\right) \times \frac{(R_U^2 - R_R^2)}{(1 - R_U^2)}$$

- An F-test gauges the extent to which restrictions reduce R^2 (as they must) by asking whether the change is small enough to be put down to sampling variance under the null
- The large-sample (asymptotic) version of the F-test is a chi-square test. Under weak assumptions like those used to derive robust standard errors, $qF \sim_a \chi^2(q)$. In other words, an F-stat times its numerator df is asymptotically (as $n \to \infty$) distributed chi-square with q df.
  - A test based on$\chi^2(q)$ is the asymptopian version of $F_{q,\, n-k-1}$. The relevant chi-square statistic is easily modified to allow for random regressors and non-normal heteroskedastic residuals. Even so, it's still more common to see an old-fashioned F statistic than a chi-square statistic in empirical publications.

3.1 Restriction Examples

- Testing additivity

$$(UR) \quad Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_{2i} + \beta_3 X_{3i} + \underbrace{\beta_{12} S_i X_{2i} + \beta_{13} S_i X_{3i} + \beta_{23} X_{2i} X_{3i}}_{2nd\ order\ terms} + \underbrace{\beta_{123} S_i X_{2i} X_{3i}}_{3rd\ order\ term} + \varepsilon_i$$

$$(R) \quad Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

In this case, an additivity null imposes four zero restrictions:

$$H_0 : \beta_{12} = \beta_{13} = \beta_{23} = \beta_{123} = 0$$

- Testing constant returns to scale (CRTS) in Cobb-Douglas production. Start with:

$$Y_i = A X_{1i}^{\beta_1} X_{2i}^{\beta_2} X_{3i}^{\beta_3} \eta_i$$

Logging:

$$(UR) \quad \ln Y_i = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \varepsilon_i, \qquad (UR)$$

where $\varepsilon_i = \ln \eta_i$. CRTS says:

$$H_0 : \beta_1 + \beta_2 + \beta_3 = 1$$

This is a single restriction on an unrestricted model with 4 parameters, so the relevant null distribution is $F_{1,\, N-4}$

  - This isn't a zero restriction
  - H_0 can be imposed by using the null to rewrite:

$$\ln Y_i = \beta_0 + (1 - \beta_2 - \beta_3) \ln X_{1i} + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \varepsilon_i$$

$$\ln Y_i - \ln X_{1i} = \beta_0 + \beta_2 (\ln X_{2i} - \ln X_{1i}) + \beta_3 (\ln X_{3i} - \ln X_{1i}) + \varepsilon_{ii}$$

$$(R) \quad Y_i^* = \beta_0 + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + \varepsilon_i$$

  - The restricted model has a new dependent variable, so the R_R^2 version of F doesn't work. But Stata's `test` command applied to the unrestricted model gets this right.

4 Testing Gender and Race Interactions in the Returns to Military Service (attached)

5 Krueger (1993): Computer Interactions in Action (attached)

```
clear
set more off
```

Data from the 2008 IPUMS-CPS http://cps.ipums.org/cps/

```
use cps_00013.dta
```

Generate usual weekly and hourly earnings

```
gen uwe =  incwage / WKSWORK1
label var uwe "Usual weekly earnings"
gen loguwe = log(uwe)
label var loguwe "log(usual weekly earnings)"
gen logahe = log(incwage/(WKSWORK1*uhrswork))
label var logahe "Log usual hourly earnings"
gen logearn = log(incwage)
label var logearn "log(annual wage and salary earnings)"
```

```
(99,260 missing values generated)
(105,868 missing values generated)
(105,868 missing values generated)
(56,097 missing values generated)
```

Generate approximate years of education

```
gen yearsEd = .
replace yearsEd=0 if educ==2
// additional labeling and output omitted from logs
label var yearsEd "Years of education (approximate)"
gen colgrad = yearsEd>=16

gen age2 = age*age
gen potex = age-yearsEd-6
label var potex "Potential experience"
gen potex2 = potex*potex
```

Code vet, mil, race, sex

```
gen military = (popstat==2)
gen veteran = (vetstat==2) | (military==1)
```

black or part black

```
gen black = (race==200 | race==801 | race==805 | race==806 | race==807)
gen blackEd = black*yearsEd
gen blackvet = black*veteran

gen female=(sex==2)
gen femEd = female*yearsEd
gen femvet = female*veteran

gen blackfem=female*black
```

```
(49,771 missing values generated)
```

keep men and women aged 30–49, include active duty, sample=inLF

```
keep if age>=30 & age<50
```

```
(146,432 observations deleted)
```

add LFvars

```
gen working = WKSWORK1>0
```

Summary Stats

```
sum age yearsEd female black veteran femvet blackvet military ///
        colgrad incwage WKSWORK1 uhrswork logahe loguwe working

sum age yearsEd female black veteran femvet blackvet military colgrad incwage
 WKSWORK1 uhrswork logahe loguwe working

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         age |     59,972    39.88245    5.681624         30         49
     yearsEd |     59,972     13.7431    2.900316          0         21
      female |     59,972    .5254786    .4993546          0          1
       black |     59,972    .1104015    .3133922          0          1
     veteran |     59,972    .0612619    .2398121          0          1
-------------+---------------------------------------------------------
      femvet |     59,972    .0090042    .0944631          0          1
    blackvet |     59,972    .0094711    .0968584          0          1
    military |     59,972    .0067698    .0820007          0          1
     colgrad |     59,972    .3235843    .4678474          0          1
     incwage |     59,972    38108.07    48959.97          0     688117
-------------+---------------------------------------------------------
    WKSWORK1 |     59,972    41.47669    19.27549          0         52
    uhrswork |     59,972    34.92955    17.70783          0         99
      logahe |     48,009    2.879673     .730158  -5.897154   9.069273
      loguwe |     48,009    6.549787    .8281364  -2.261763   11.15625
     working |     59,972     .852131     .354973          0          1
```

Additive model

```
reg logahe yearsEd veteran black female


      Source |       SS           df       MS      Number of obs   =    48,009
-------------+----------------------------------   F(4, 48004)     =   3026.50
       Model |  5154.67766         4  1288.66941   Prob > F        =    0.0000
    Residual |  20439.8621    48,004  .425794977   R-squared       =    0.2014
-------------+----------------------------------   Adj R-squared   =    0.2013
       Total |  25594.5397    48,008  .533130723   Root MSE        =    .65253

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |    .104886   .0010596    98.98   0.000     .1028091    .1069629
     veteran |   .0367164   .0121419     3.02   0.002      .012918    .0605147
       black |  -.0988435    .009582   -10.32   0.000    -.1176244   -.0800627
      female |  -.2972256   .0060911   -48.80   0.000    -.3091643   -.2852869
       _cons |   1.570199   .0152843   102.73   0.000     1.540242    1.600157
------------------------------------------------------------------------------
```

Regression of wages on eduation and veteran status, by sex, control for race

```
bys female: reg logahe yearsEd veteran black


-------------------------------------------------------------------------------
-> female = 0

      Source |       SS           df       MS      Number of obs   =    24,552
-------------+----------------------------------   F(3, 24548)     =   1844.86
       Model |  2348.76452         3  782.921508   Prob > F        =    0.0000
    Residual |  10417.6685    24,548  .424379524   R-squared       =    0.1840
-------------+----------------------------------   Adj R-squared   =    0.1839
       Total |  12766.4331    24,551  .519996459   Root MSE        =    .65144

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1024476   .0014073    72.80   0.000     .0996893    .1052059
     veteran |   .0348462   .0131176     2.66   0.008     .0091349    .0605575
       black |  -.1908072   .0143528   -13.29   0.000    -.2189395   -.1626749
       _cons |    1.61261   .0199472    80.84   0.000     1.573512    1.651707
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> female = 1

      Source |       SS           df       MS      Number of obs   =    23,457
-------------+----------------------------------   F(3, 23453)     =   1523.13
       Model |  1945.86697         3  648.622324   Prob > F        =    0.0000
    Residual |  9987.43719    23,453  .425849025   R-squared       =    0.1631
-------------+----------------------------------   Adj R-squared   =    0.1630
       Total |  11933.3042    23,456  .508752735   Root MSE        =    .65257

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1081747   .0016071    67.31   0.000     .1050246    .1113248
     veteran |   .0681413   .0318518     2.14   0.032     .0057097    .1305729
       black |  -.0251327   .0128612    -1.95   0.051    -.0503416    .0000761
       _cons |   1.216732   .0231711    52.51   0.000     1.171315    1.262149
------------------------------------------------------------------------------
```

Regression of wages on eduation and veteran status interacted with sex

```
reg logahe yearsEd femEd veteran femvet black female blackfem


      Source |       SS           df       MS      Number of obs   =    48,009
-------------+----------------------------------   F(7, 48001)     =   1743.95
       Model |  5189.43401         7  741.347715   Prob > F        =    0.0000
    Residual |  20405.1057    48,001  .425097513   R-squared       =    0.2028
-------------+----------------------------------   Adj R-squared   =    0.2026
       Total |  25594.5397    48,008  .533130723   Root MSE        =      .652

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1024476   .0014084    72.74   0.000      .099687    .1052082
       femEd |   .0057271   .0021359     2.68   0.007     .0015407    .0099135
     veteran |   .0348462   .0131287     2.65   0.008     .0091138    .0605786
      femvet |   .0332951   .0344254     0.97   0.333    -.0341792    .1007694
       black |  -.1908072   .0143649   -13.28   0.000    -.2189626   -.1626518
      female |  -.3958777   .0305699   -12.95   0.000    -.4557951   -.3359604
    blackfem |   .1656745   .0192735     8.60   0.000     .1278981    .2034508
       _cons |    1.61261   .0199641    80.78   0.000      1.57348     1.65174
------------------------------------------------------------------------------
```

Test interactions jointly

```
test femEd femvet
test femEd femvet blackfem
lincom black+blackfem


 ( 1)  femEd = 0
 ( 2)  femvet = 0

       F(  2, 48001) =    4.14
            Prob > F =    0.0160

 ( 1)  femEd = 0
 ( 2)  femvet = 0
 ( 3)  blackfem = 0

       F(  3, 48001) =   27.25
            Prob > F =    0.0000

 ( 1)  black + blackfem = 0

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0251327   .0128499    -1.96   0.050    -.0503186    .0000532
------------------------------------------------------------------------------
```

lincom estimates linear combinations

Regression of wages on education and veteran status, by race, control for sex

```
bys black: reg logahe yearsEd veteran female


-------------------------------------------------------------------------------
-> black = 0

      Source |       SS           df       MS      Number of obs   =    42,771
-------------+----------------------------------   F(3, 42767)     =   3644.61
       Model |  4732.42079         3  1577.4736   Prob > F        =    0.0000
    Residual |  18510.5762    42,767  .432823818   R-squared       =    0.2036
-------------+----------------------------------   Adj R-squared   =    0.2036
       Total |   23242.997    42,770  .543441595   Root MSE        =    .65789

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1048038     .00111    94.41   0.000     .1026281    .1069795
     veteran |   .0245867   .0132534     1.86   0.064    -.0013903    .0505637
      female |  -.3160608   .0064956   -48.66   0.000    -.3287923   -.3033293
       _cons |   1.581149   .0159949    98.85   0.000     1.549799      1.6125
------------------------------------------------------------------------------

-------------------------------------------------------------------------------
-> black = 1

      Source |       SS           df       MS      Number of obs   =     5,238
-------------+----------------------------------   F(3, 5234)      =    325.53
       Model |  353.444082         3  117.814694   Prob > F        =    0.0000
    Residual |  1894.24545     5,234  .361911627   R-squared       =    0.1572
-------------+----------------------------------   Adj R-squared   =    0.1568
       Total |  2247.68954     5,237  .429194107   Root MSE        =    .60159

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1066224   .0035997    29.62   0.000     .0995655    .1136792
     veteran |   .1253284   .0292934     4.28   0.000     .0679011    .1827557
      female |  -.1386085   .0172364    -8.04   0.000    -.1723989    -.104818
       _cons |   1.349591   .0506443    26.65   0.000     1.250307    1.448875
------------------------------------------------------------------------------
```

Regression of wages on education and veteran status interacted with race

```
reg logahe yearsEd blackEd veteran blackvet black female blackfem


      Source |       SS           df       MS      Number of obs   =    48,009
-------------+----------------------------------   F(7, 48001)     =   1744.07
       Model |  5189.71807         7  741.388295   Prob > F        =    0.0000
    Residual |  20404.8217    48,001  .425091595   R-squared       =    0.2028
-------------+----------------------------------   Adj R-squared   =    0.2027
       Total |  25594.5397    48,008  .533130723   Root MSE        =    .65199

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsEd |   .1048038   .0011001    95.27   0.000     .1026476      .10696
     blackEd |   .0018186   .0040534     0.45   0.654    -.0061261    .0097633
     veteran |   .0245867   .0131345     1.87   0.061    -.0011571    .0503306
    blackvet |   .1007417   .0343572     2.93   0.003      .033401    .1680823
       black |  -.2315583   .0571302    -4.05   0.000    -.3435343   -.1195824
      female |  -.3160608   .0064373   -49.10   0.000    -.3286781   -.3034436
    blackfem |   .1774524   .0197584     8.98   0.000     .1387256    .2161792
       _cons |   1.581149   .0158514    99.75   0.000     1.550081    1.612218
------------------------------------------------------------------------------
```

Test interactions jointly

```
test blackEd blackvet
test blackEd blackvet blackfem
lincom black+blackfem


 ( 1)  blackEd = 0
 ( 2)  blackvet = 0

       F(  2, 48001) =    4.47
            Prob > F =    0.0114

 ( 1)  blackEd = 0
 ( 2)  blackvet = 0
 ( 3)  blackfem = 0

       F(  3, 48001) =   27.48
            Prob > F =    0.0000

 ( 1)  black + blackfem = 0

------------------------------------------------------------------------------
      logahe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   -.054106   .0574678    -0.94   0.346    -.1667436    .0585316
------------------------------------------------------------------------------
```

14.32
SP2023

Krueger 1993

TABLE I
PERCENT OF WORKERS IN VARIOUS CATEGORIES WHO DIRECTLY USE A COMPUTER AT WORK

Group	1984	1989
All workers	24.6	37.4
Gender		
Men	21.2	32.3
Women	29.0	43.4
Education		
Less than high school	5.0	7.8
High school	19.3	29.3
Some college	30.6	45.3
College	41.6	58.2
Postcollege	42.8	59.7
Race		
White	25.3	38.5
Black	19.4	27.7
Age		
Age 18–24	19.7	29.4
Age 25–39	29.2	41.5
Age 40–54	23.6	39.1
Age 55–65	16.9	26.3
Occupation		
Blue-collar	7.1	11.6
White-collar	33.0	48.4
Union status		
Union member	20.2	32.5
Nonunion	28.0	41.1
Hours		
Part-time	23.7	36.3
Full-time	28.9	42.7
Region		
Northeast	25.5	38.0
Midwest	23.4	36.0
South	23.2	36.5
West	27.0	39.9

Source. Author's tabulations of the 1984 and 1989 October Current Population Surveys. The sample size is 61,712 for 1984 and 62,748 for 1989.

TABLE II
OLS REGRESSION ESTIMATES OF THE EFFECT OF COMPUTER USE ON PAY
(DEPENDENT VARIABLE: ln (HOURLY WAGE))

Independent variable	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.937	0.750	0.928	2.086	0.905	1.094
	(0.005)	(0.023)	(0.026)	(0.006)	(0.024)	(0.026)
Uses computer at work (1 = yes)	0.276	0.170	0.140	0.325	0.188	0.162
	(0.010)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Years of education	—	0.069	0.048	—	0.075	0.055
		(0.001)	(0.002)		(0.002)	(0.002)
Experience	—	0.027	0.025	—	0.027	0.025
		(0.001)	(0.001)		(0.001)	(0.001)
Experience-squared ÷ 100	—	−0.041	−0.040	—	−0.041	−0.040
		(0.002)	(0.002)		(0.002)	(0.002)
Black (1 = yes)	—	−0.098	−0.066	—	−0.121	−0.092
		(0.013)	(0.012)		(0.013)	(0.012)
Other race (1 = yes)	—	−0.105	−0.079	—	−0.029	−0.015
		(0.020)	(0.019)		(0.020)	(0.020)
Part-time (1 = yes)	—	−0.256	−0.216	—	−0.221	−0.183
		(0.010)	(0.010)		(0.010)	(0.010)
Lives in SMSA (1 = yes)	—	0.111	0.105	—	0.138	0.130
		(0.007)	(0.007)		(0.007)	(0.007)
Veteran (1 = yes)	—	0.038	0.041	—	0.025	0.031
		(0.011)	(0.011)		(0.012)	(0.011)
Female (1 = yes)	—	−0.162	−0.135	—	−0.172	−0.151
		(0.012)	(0.012)		(0.012)	(0.012)
Married (1 = yes)	—	0.156	0.129	—	0.159	0.143
		(0.011)	(0.011)		(0.011)	(0.011)
Married*Female	—	−0.168	−0.151	—	−0.141	−0.131
		(0.015)	(0.015)		(0.015)	(0.015)
Union member (1 = yes)	—	0.181	0.194	—	0.182	0.189
		(0.009)	(0.009)		(0.010)	(0.010)
8 Occupation dummies	No	No	Yes	No	No	Yes
R^2	0.051	0.446	0.491	0.082	0.451	0.486

Notes. Standard errors are shown in parentheses. Sample size is 13,335 for 1984 and 13,379 for 1989. Columns (2), (3), (5), and (6) also include three region dummy variables.

TABLE III

The Return to Various Uses of Computers, October 1989[a]

(Dependent Variable: ln (hourly wage))

Use of computer at work	Proportion	Coefficient (std. error)
Uses computer at work for any task[b]	0.398	0.145
		(0.010)
Specific Task[c]		
Word processing	0.165	0.017
		(0.012)
Bookkeeping	0.100	−0.058
		(0.013)
Computer-assisted design	0.039	0.026
		(0.020)
Electronic mail	0.063	0.149
		(0.016)
Inventory control	0.102	−0.056
		(0.013)
Programming	0.077	0.052
		(0.031)
Desktop publishing or newsletters	0.036	−0.047
		(0.021)
Spread sheets	0.094	0.079
		(0.015)
Sales	0.060	−0.002
		(0.016)
Computer games	0.019	−0.109
		(0.026)
R^2		0.495

a. The sample and other explanatory variables are the same as in column (6) of Table II.

b. The computer use dummy variable equals one if the worker uses computers for any of the ten enumerated tasks or for any other task.

c. The dummy variables for any specific computer task, and the dummy variable for any computer use, are not mutually exclusive.

TABLE IV

THE RETURN TO COMPUTER USE AT WORK, HOME, AND WORK AND HOME

(STANDARD ERRORS ARE SHOWN IN PARENTHESES.)

Type of computer use	October 1984 (1)	October 1989 (2)	Percent of sample, 1989 (3)
Uses computer at work	0.165	0.177	39.8
	(0.009)	(0.009)	
Uses computer at home	0.056	0.070	12.5
	(0.021)	(0.019)	
Uses computer at home and work	0.006	0.017	8.6
	(0.029)	(0.023)	
Sample size	13,335	13,379	

Notes. The table reports coefficients for three dummy variables estimated from log hourly wage regressions. The other explanatory variables in the regressions are education, experience and its square, two race dummies, three region dummies, dummy variables indicating part-time status, residence in an SMSA, veteran status, gender, marital status, union membership, and an interaction between marital status and gender. Covariates are the same as in columns (2) and (5) of Table II.

TABLE VII

OLS REGRESSION ESTIMATES OF THE EFFECT OF COMPUTER USE ON PAY

(DEPENDENT VARIABLE: ln (HOURLY WAGE))

Independent variable	October 1984			October 1989		
	(1)	(2)	(3)	(4)	(5)	(6)
Uses computer at work (1 = yes)	—	0.170	0.073	—	0.188	0.005
		(0.008)	(0.048)		(0.008)	(0.043)
Computer use*Education	—	—	0.007	—	—	0.013
			(0.003)			(0.003)
Years of education	0.076	0.069	0.067	0.086	0.075	0.071
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Experience	0.027	0.027	0.027	0.027	0.027	0.027
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience-squared ÷ 100	−0.042	−0.041	−0.042	−0.044	−0.041	−0.042
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Black (1 = yes)	−0.106	−0.098	−0.099	−0.141	−0.121	−0.122
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Other race (1 = yes)	−0.120	−0.105	−0.106	−0.037	−0.029	−0.032
	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)
Part-time (1 = yes)	−0.287	−0.256	−0.256	−0.261	−0.221	−0.221
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Lives in SMSA (1 = yes)	0.123	0.111	0.111	0.148	0.138	0.138
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Veteran (1 = yes)	0.043	0.038	0.039	0.027	0.025	0.029
	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.012)
Female (1 = yes)	−0.140	−0.162	−0.160	−0.142	−0.172	−0.168
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Married (1 = yes)	0.162	0.156	0.156	0.169	0.159	0.158
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Married*Female	−0.171	−0.168	−0.168	−0.146	−0.141	−0.139
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Union member (1 = yes)	0.167	0.181	0.181	0.164	0.182	0.182
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
R^2	0.429	0.446	0.446	0.428	0.451	0.452
Mean-squared error	0.168	0.163	0.163	0.176	0.169	0.169

Notes. Standard errors are shown in parentheses. Sample size is 13,335 for 1984 and 13,379 for 1989. Regressions also include three region dummy variables and an intercept.

DiNardo-Pischke 2007

TABLE I
PERCENT OF WORKERS IN VARIOUS CATEGORIES WHO USE DIFFERENT TOOLS ON THEIR JOB

Group	U. S. 1984	U. S. 1989	U. S. 1993	Germany 1979	Germany 1985–1986	Germany 1991–1992
	Percentage that are computer users					
All workers	25.1	37.4	46.6	8.5	18.5	35.3
Men	21.6	32.2	41.1	7.9	18.5	36.4
Women	29.6	43.8	53.2	9.7	18.5	33.5
Less than high school	5.1	7.7	10.4	3.2	4.3	9.9
High school	19.2	28.4	34.6	8.5	18.3	32.7
Some college	30.6	45.0	53.1	8.5	24.8	48.4
College	42.4	58.8	70.2	13.4	30.5	61.6
Age 18–24	20.5	29.6	34.3	10.1	13.8	27.8
Age 25–39	29.6	41.4	49.8	9.6	21.6	39.9
Age 40–54	23.9	38.9	50.0	6.6	17.2	35.9
Age 55–64	17.7	27.0	37.3	5.9	13.5	23.7
Blue-collar	7.1	11.2	56.6	1.2	3.5	10.7
White-collar	39.7	56.6	67.6	12.8	28.9	50.2
Part-time	14.8	24.4	29.3	6.4	14.7	26.5
Full-time	29.3	42.3	51.0	8.7	19.1	37.0
	Percentage of all workers who use a specific tool					
Computer	25.1	37.4	46.6	8.5	18.5	35.3
Calculator				19.6	35.7	44.2
Telephone				41.8	43.7	58.4
Pen/pencil				54.9	53.4	65.6
Work while sitting[a]				30.8	19.3	—
Hand tool (e.g., hammer)				29.4	32.9	30.5
Number of obs.	61,704	62,748	59,852	19,427	22,353	20,042

a. Variable definition differs in 1979 and 1985–1986. In 1979 it refers to "Never or rarely standing," and in 1985–1986 it refers to "Often or almost always sitting."

Columns 1 to 3 are from Table 3 in Autor, Katz, and Krueger [1996] and come from the October *Current Population Survey.* German data are from the *Qualification and Career Survey.*

resulting approximation should be rather good because of the large number of brackets. Adopting a similar specification, we find that this earnings variable yields the same return to schooling in 1985–1986 as reported by Krueger and Pischke [1995] with a continuous earnings variable for 1988. Years of education are imputed from information on schools attended and degrees obtained following Krueger and Pischke [1995]. When we bracket the earnings variable in the October 1984 CPS to be comparable with our 1979 data, in a regression similar to Krueger's we find a computer coefficient of 0.1697 using the original wage variable, and 0.1701 using the bracketed variable. Standard errors are about 3 percent larger with the bracketed variable.

TABLE II
OLS REGRESSIONS FOR THE EFFECT OF COMPUTER USE ON PAY
DEPENDENT VARIABLE: LOG HOURLY WAGE
(STANDARD ERRORS IN PARENTHESES)

Independent variable	U. S. 1984	U. S. 1989	U. S. 1993	Germany 1979	Germany 1985–1986	Germany 1991–1992
Computer	0.171	0.188	0.204	0.112	0.157	0.171
	(0.008)	(0.008)	(0.008)	(0.010)	(0.007)	(0.006)
Years of schooling	0.068	0.075	0.081	0.073	0.063	0.072
	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Experience	0.028	0.028	0.026	0.030	0.035	0.030
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Experience^2/100	−0.043	−0.043	−0.041	−0.052	−0.058	−0.046
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
R^2	0.444	0.448	0.424	0.267	0.280	0.336
Number of obs.	13,335	13,379	13,305	19,427	22,353	20,042

Columns 1 to 3 are from Table 4 in Autor, Katz, and Krueger [1996]. Data for columns 1 to 3 are from the October *Current Population Survey;* data for columns 4 to 6 are from the *Qualification and Career Survey.* All models also include an intercept, a dummy for part-time, large city/SMSA status, female, married, female*married. Regressions for the United States in columns 1 to 3 also include dummies for black, other race, veteran status, union membership, and three regions. Regressions for Germany in columns 4 to 6 also include a dummy for civil servants (*Beamter*).

TABLE III

OLS Regression for the Effect of Different Tools on Pay

Dependent Variable: Log Hourly Wage

(standard errors in parentheses)

Independent variable	Germany 1979	Germany 1985–86	Germany 1991–92	Germany 1979	Germany 1979	Germany 1985–1986	Germany 1991–1992
Occupation indicators	No	No	No	501	501	742	1071
Grades and father's Occupation[a]	No	No	No	No	Yes	No	No
				Tools entered separately			
Computer	0.112	0.157	0.171	0.025	0.022	0.076	0.083
	(0.010)	(0.007)	(0.006)	(0.011)	(0.011)	(0.008)	(0.007)
Calculator	0.087	0.128	0.129	0.027	0.025	0.061	0.054
	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)	(0.007)	(0.006)
Telephone	0.131	0.114	0.136	0.060	0.057	0.059	0.072
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Pen/pencil	0.123	0.112	0.127	0.055	0.052	0.055	0.050
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Work while sitting	0.106	0.101	—	0.042	0.041	0.036	—
	(0.006)	(0.007)		(0.008)	(0.008)	(0.008)	
Hand tool (e.g., hammer)	−0.117	−0.086	−0.091	−0.048	−0.045	−0.020	−0.020
	(0.007)	(0.006)	(0.006)	(0.009)	(0.009)	(0.008)	(0.008)

	Tools entered together						
Computer	0.066	0.105	0.126	0.027	0.024	0.067	0.069
	(0.010)	(0.008)	(0.007)	(0.011)	(0.011)	(0.008)	(0.007)
Calculator	0.017	0.053	0.044	0.015	0.014	0.032	0.022
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)
Telephone	0.072	0.043	0.045	0.043	0.041	0.035	0.048
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Pen/pencil	0.062	0.031	0.035	0.040	0.038	0.024	0.007
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Work while sitting	0.058	0.050	—	0.036	0.035	0.032	—
	(0.007)	(0.007)		(0.008)	(0.008)	(0.008)	

a. Two variables for self-reported grades in math and German and eleven dummy variables for father's education.

Data are from the *Qualification and Career Survey*. All regressions also include an intercept, years of schooling, experience and experience squared, dummies for part-time, city, female, married, married*female, and for civil servants (*Beamter*).