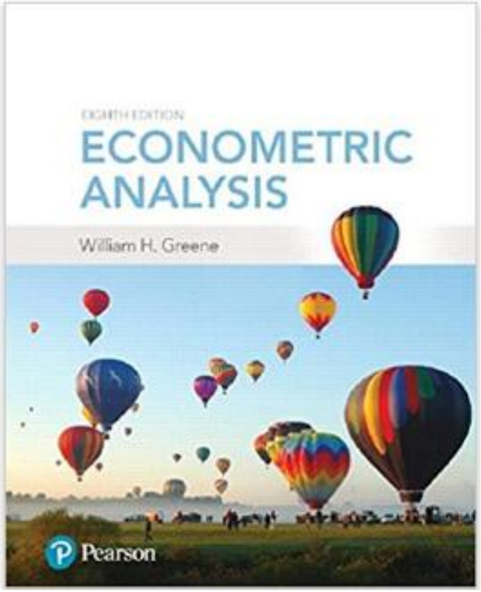
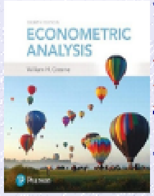
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



http://people.stern.nyu.edu/wgreene/Econometrics/Econometrics.htm

- Course Outline
- Class Notes
- Problem Sets
- Data Sets
- Text
- Solutions for 6th Edition
- Queries



Applied Econometrics

Professor William Greene

This is an intermediate level, Ph.D. course in Applied Econometrics. Topics to be studied include specification, estimation, and inference in the context of models that include then extend beyond the standard linear multiple regression framework. After a review of the linear model, we will develop the asymptotic distribution theory necessary for analysis of generalized linear and nonlinear models. We will then turn to instrumental variables, maximum likelihood, generalized method of moments (GMM), and two step estimation methods. Inference techniques used in the linear regression framework such as t and F tests will be extended to include Wald, Lagrange multiplier and likelihood ratio and tests for nonnested hypotheses such as the Hausman specification test. Specific modelling frameworks will include the linear regression model and extensions to models for panel data, multiple equation models, and models for discrete choice. The discussion will include elements of the following topics:

The Paradigm of Econometrics
Classical Linear Regression Model. Part 1. Specification and Computation
Classical Linear Regression Model. Part 2. Statistical Inference in Finite Samples
Asymptotic Theory and Instrumental Variables Estimation
Nonlinear Regression Models
The Generalized Regression Model
Methods of Estimation

- Instrumental variables estimation
- Maximum likelihood estimation
- Generalized method of moments (GMM) estimation
- Two step estimation
- Estimation by simulation; models with unobserved heterogeneity
- Bayesian methods in econometrics
- Non and semiparametric estimators

Abstract

Overview: This is an intermediate level, Ph.D. course in Applied Econometrics. Topics to be studied include specification, estimation, and inference in the context of models that include then extend beyond the standard linear multiple regression framework. After a review of the linear model, we will develop the theory necessary for analysis of generalized linear and nonlinear models.

Topics to be examined:

- Regression modeling
- Instrumental variables
- Robust estimation and inference
- Causal inference
- Nonlinear modeling
- Cross section, time series and panel data

Objective:

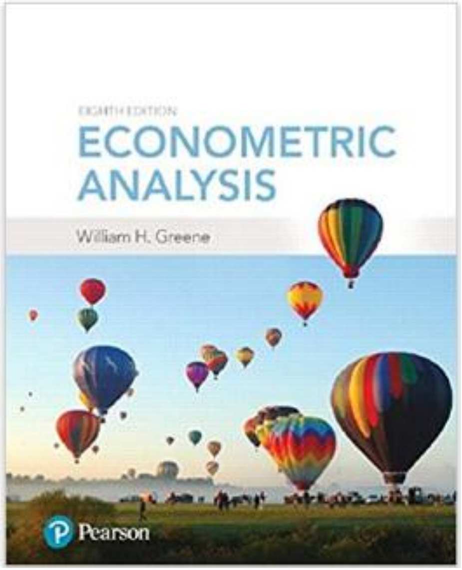
Preparation to read and carry out empirical social science research using modern econometric methods.

Prerequisites

- A previous course that used linear regression
- Mathematical statistics
- Matrix algebra

Readings

- Main text: Greene, W., Econometric Analysis, 8th Edition, Prentice Hall, 2017.



- A few articles
- Notes and materials on the course website:

http://people.stern.nyu.edu/wgreene/Econometrics/Econometrics.htm

Course Schedule

No class on:

Thursday, September 21
***Tuesday, November 21**
Thursday, November 23
***Tuesday, November 28**

Added class on:

Tuesday, December 12

Midterm: October 24

Final: Take home; posted December 7, due December 19

Econometrics I: Applied Econometrics

Stern School of Business

Professor W. Greene
Department of Economics
Office:;MEC 7-90, Ph. 998-0876
e-mail: wgreene@stern.nyu.edu
WWW: http://people.stern.nyu.edu/wgreene

Abstract: This is an intermediate level, Ph.D. course in *Applied Econometrics*. Topics to be studied include specification, estimation, and inference in the context of models that include then extend beyond the standard linear multiple regression framework. After a review of the linear model, we will develop the asymptotic distribution theory necessary for robust estimation and inference and analysis of linear and nonlinear models. We will then turn to instrumental variables, maximum likelihood, generalized method of moments (GMM), and two step estimation methods. Inference techniques used in the linear regression framework such as t and F tests will be extended to include Wald, Lagrange multiplier and likelihood ratio and tests for nonnested hypotheses such as the Hausman specification test. Specific modelling frameworks will include the linear regression model and extensions to models for panel data, multiple equation models, time series models and models for discrete choice and sample selection.

Prerequisites: Multivariate calculus, matrix algebra, probability and distribution theory, statistical inference, and an introduction to the multiple linear regression model. Appendices A and B in *Greene (2017)* are assumed. We will survey the parts of Appendix C that would have appeared in prerequisite courses. A significant part of this course will focus on the advanced parts of Appendices C and D. We will also make use of a few of the results in Appendix E (optimization).

Practicals

- Software
  - NLOGIT provided, supported
    http://people.stern.nyu.edu/wgreene/Econometrics/NLsetup.exe
  - SAS, Stata, EViews optional, your choice
  - R, Matlab, Gauss, others
  - Questions and review as requested
- Problem Sets: (more details later)

Course Requirements

- Problem sets: 5 (30%)
- Midterm, in class (30%)
- Final exam, take home (40%)

Econometrics I

Part 1 - Paradigm

Econometrics: Paradigm

- Theoretical foundations
  - Behavioral modeling: Optimization, labor supply, demand equations, etc.
  - Microeconometrics and macroeconometrics
- Mathematical Elements
- Statistical foundations
- 'Model' building – the econometric model
  - Mathematical elements
  - The underlying truth – is there one?

Objectives

- Understanding covariation:
  - **What is the income elasticity of the demand for health care in an economy?**
- Understanding a relationship:
  - **How will drivers respond to the availability of a new toll road?**
- Causal Inference: The search for "causal" effects
  - **Did an antitrust intervention that broke up a British boarding schools cartel have an impact on fees charged?**
- Prediction of an outcome of interest
  - **Does internet "buzz" help to predict movie success?**

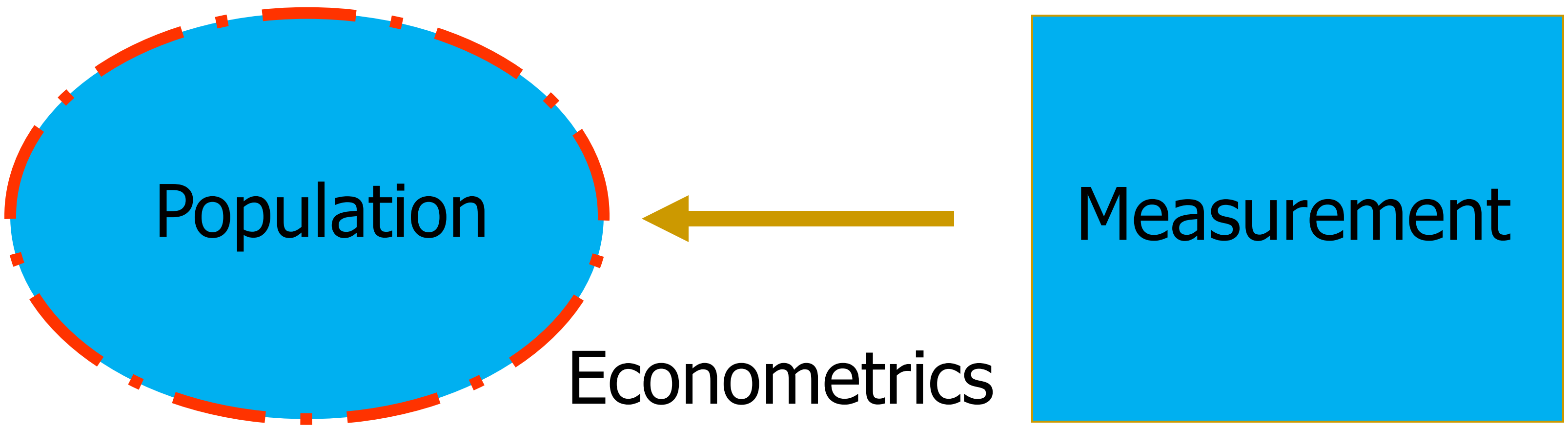
Trends in Econometrics

- Small structural models
- Pervasiveness of an econometrics paradigm
- Non- and semiparametric methods vs. parametric
- Robust methods / Estimation and inference
- Nonlinear modeling (the role of software)
- Behavioral and structural modeling vs. “reduced form,” “covariance analysis”
- Identification and “causal” effects

Data Structures

- Observation mechanisms
  - Passive, nonexperimental (the usual)
  - Randomly assigned experiment (wishful)
  - Active, experimental (occasional)
  - The ‘natural experiment’ (occasional, limited)
- Data types
  - Cross section
  - Pure time series
  - Panel –
    - Longitudinal data – NLS
    - Country macro data – Penn W.T.
  - Financial data

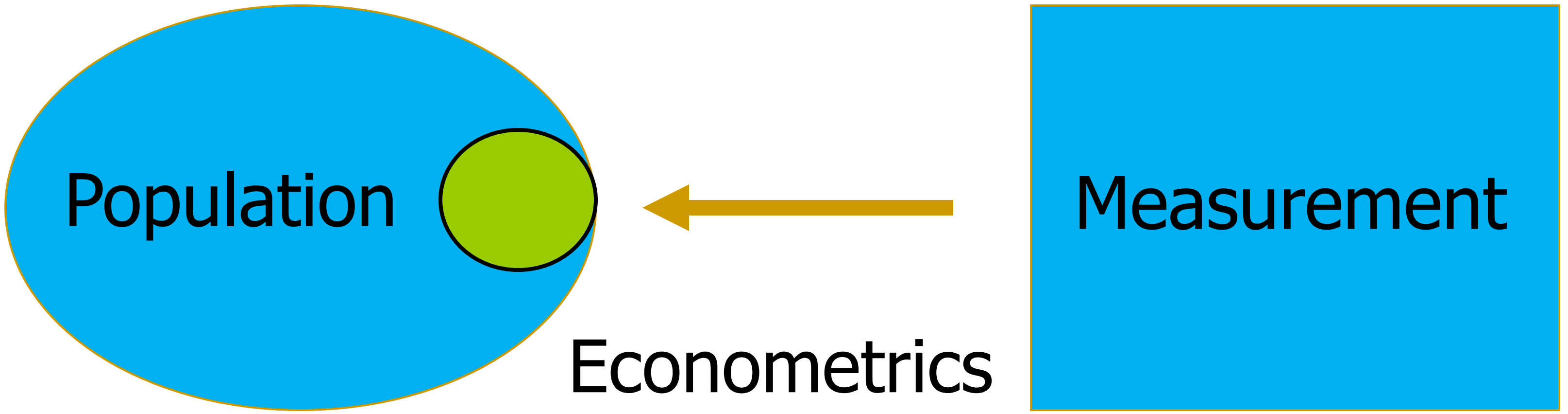
Paradigm: Classical Inference



Characteristics
Behavior Patterns
Choices

Imprecise inference about the entire population – sampling theory and asymptotics

Paradigm: Bayesian Inference



Characteristics
Behavior Patterns
Choices

Sharp, ‘exact’ inference about only the sample – the ‘posterior’ density.

An Application: Cornwell and Rupert Labor Market Data Is Wage Related to Education?

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are

EXP	= work experience
WKS	= weeks worked
OCC	= occupation, 1 if blue collar,
IND	= 1 if manufacturing industry
SOUTH	= 1 if resides in south
SMSA	= 1 if resides in a city (SMSA)
MS	= 1 if married
FEM	= 1 if female
UNION	= 1 if wage set by union contract
ED	= years of education
LWAGE	= log of wage = dependent variable in regressions

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," *Journal of Applied Econometrics*, 3, 1988, pp. 149-155.

Data Editor

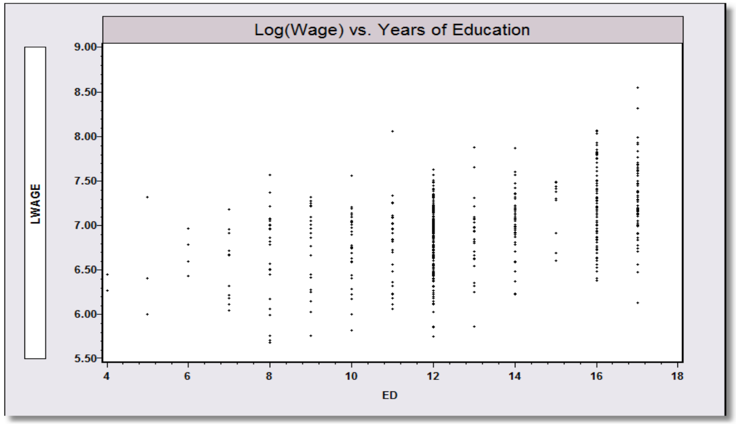
28/900 Vars; 11111 Rows: 4165 Obs Cell: 0

	LOGWAGE	EDUC
1 »	5.56068	9
2 »	5.72031	9
3 »	5.99645	9
4 »	5.99645	9
5 »	6.06146	9
6 »	6.17379	9
7 »	6.24417	9
8 »	6.16331	11
9 »	6.21461	11
10 »	6.2634	11
11 »	6.54391	11
12 »	6.69703	11
13 »	6.79122	11
14 »	6.81564	11
15 »	5.65249	12
16 »	6.43615	12
17 »	6.54822	12
18 »	6.60259	12
19 »	6.6958	12
20 »	6.77878	12
21 »	6.86066	12
22 »	[illegible]	[illegible]

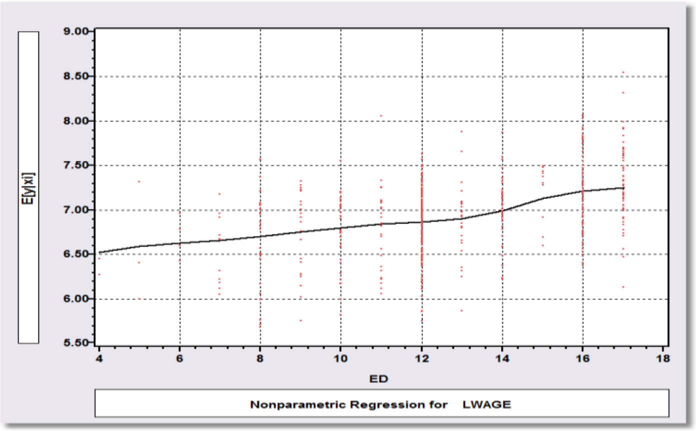
Model Building in Econometrics

- Role of the assumptions
- Parameterizing the model
  - Nonparametric analysis
  - Semiparametric analysis
  - Parametric analysis
- Sharpness of inferences

Application: Is there a relationship between (log) Wage and Education?

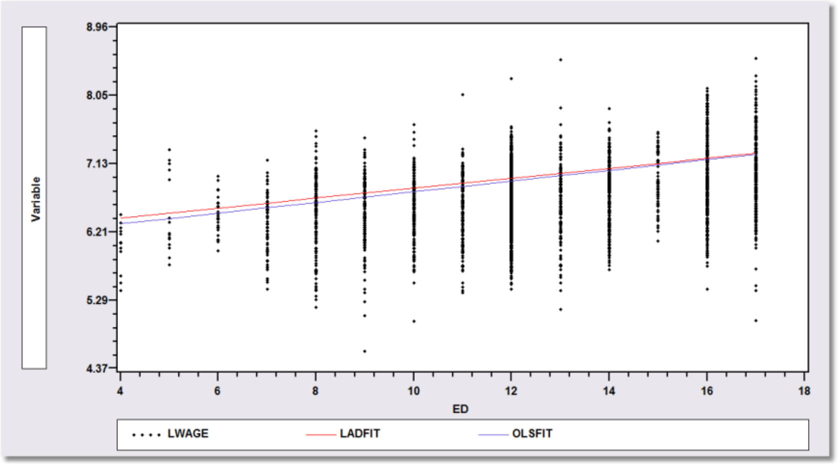


Nonparametric Regression
Kernel regression of log wage on education



Semiparametric Regression:
Least absolute deviations regression

Parametric Regression: Least squares – maximum likelihood – regression



A First Look at the Data
Descriptive Statistics

- Basic Measures of Location and Dispersion
- Graphical Devices
  - Box Plots
  - Histogram
  - Kernel Density Estimator

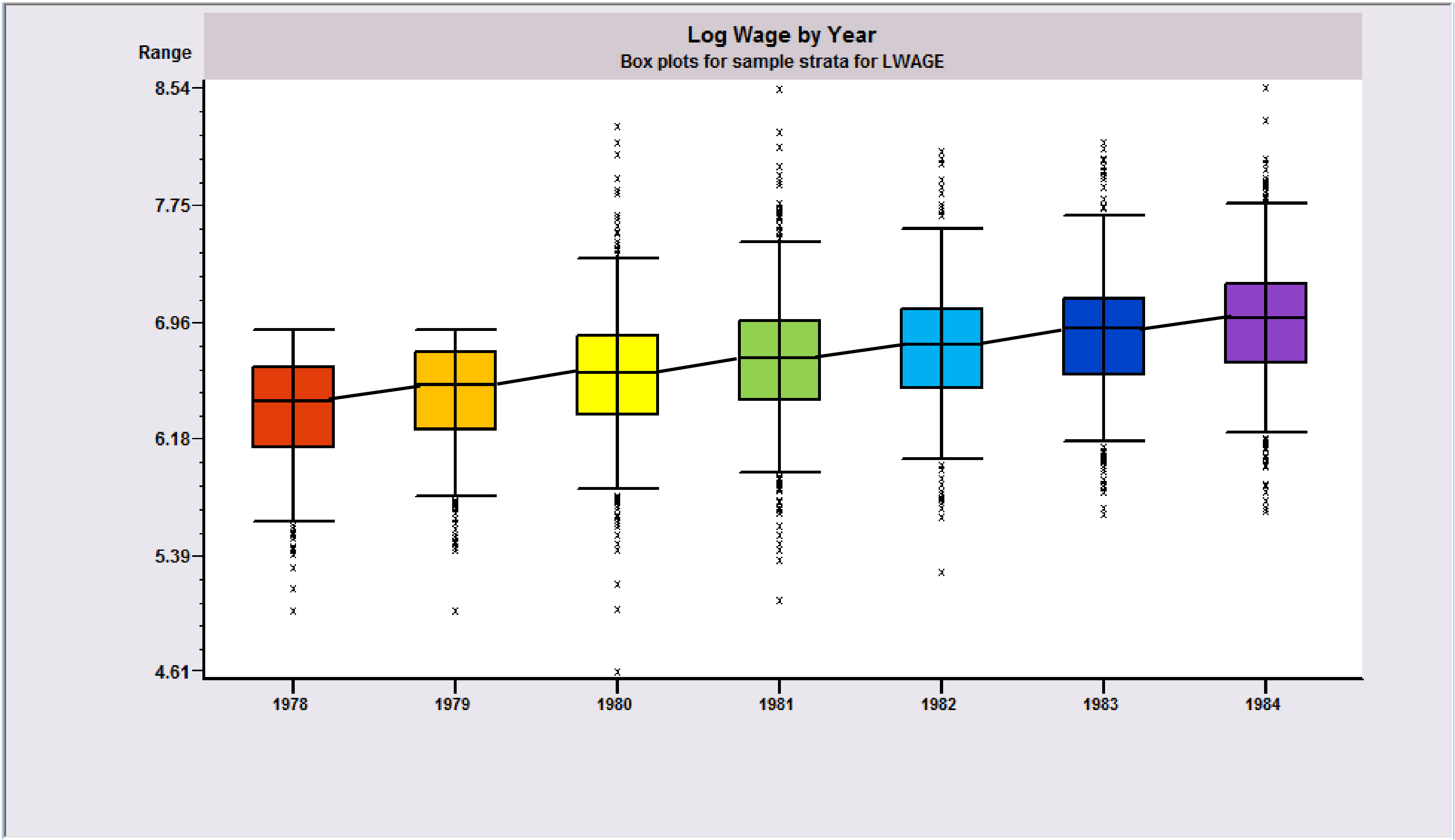
Descriptive Statistics for 11 variables

Variable	Mean	Std.Dev.	Minimum	Maximum	Cases	Missing
EXP	19.85378	10.96637	1.0	51.0	4165	0
WKS	46.81152	5.129098	5.0	52.0	4165	0
OCC	.511164	.499935	0.0	1.0	4165	0
IND	.395438	.489003	0.0	1.0	4165	0
SOUTH	.290276	.453944	0.0	1.0	4165	0
SMSA	.653782	.475821	0.0	1.0	4165	0
MS	.814406	.388826	0.0	1.0	4165	0
FEM	.112605	.316147	0.0	1.0	4165	0
UNION	.363986	.481202	0.0	1.0	4165	0
LWAGE	6.676346	.461512	4.605170	8.537000	4165	0
YEAR	4.0	2.000240	1.0	7.0	4165	0

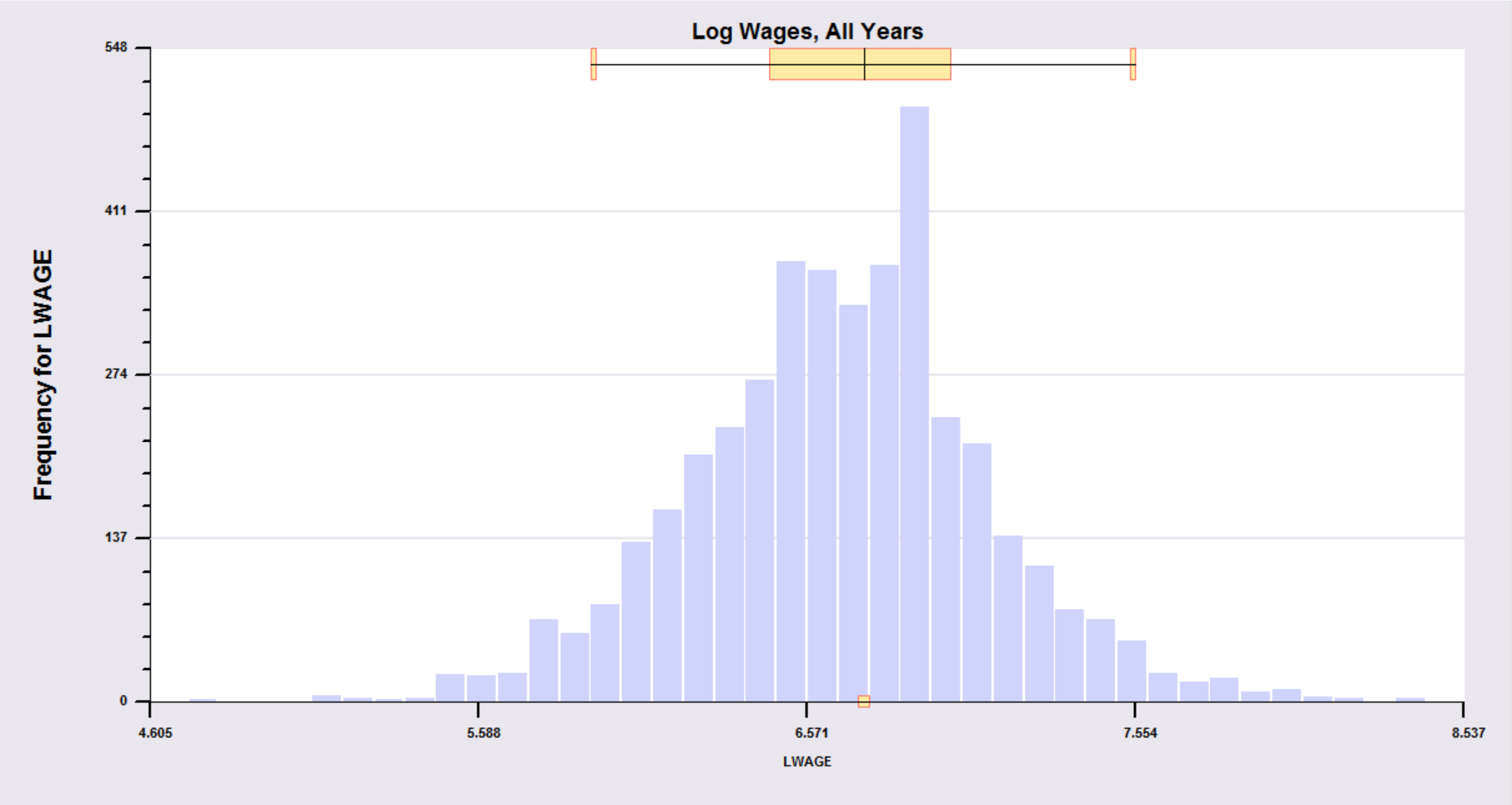
Descriptive Statistics for LWAGE
Stratification is based on YEAR

Subsample	Mean	Std.Dev.	Cases	Sum of wts	Missing
YEAR = 1	6.375173	.388426	595	595.00	0
YEAR = 2	6.465212	.362702	595	595.00	0
YEAR = 3	6.596717	.446691	595	595.00	0
YEAR = 4	6.696079	.440750	595	595.00	0
YEAR = 5	6.786454	.424013	595	595.00	0
YEAR = 6	6.864045	.424021	595	595.00	0
YEAR = 7	6.950745	.438403	595	595.00	0
Full Sample	6.676346	.461512	4165	4165.00	0

Box Plots
Shows upward trend in median log wage



Histogram: Pooled data obscure within person variation



What is the source of the variance, variation across people or variation over time?

Kernel density estimator suggests the underlying distribution for a continuous variable

$$\hat{f}(x_m^*) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{B}K\left[\frac{x_i - x_m^*}{B}\right]$$

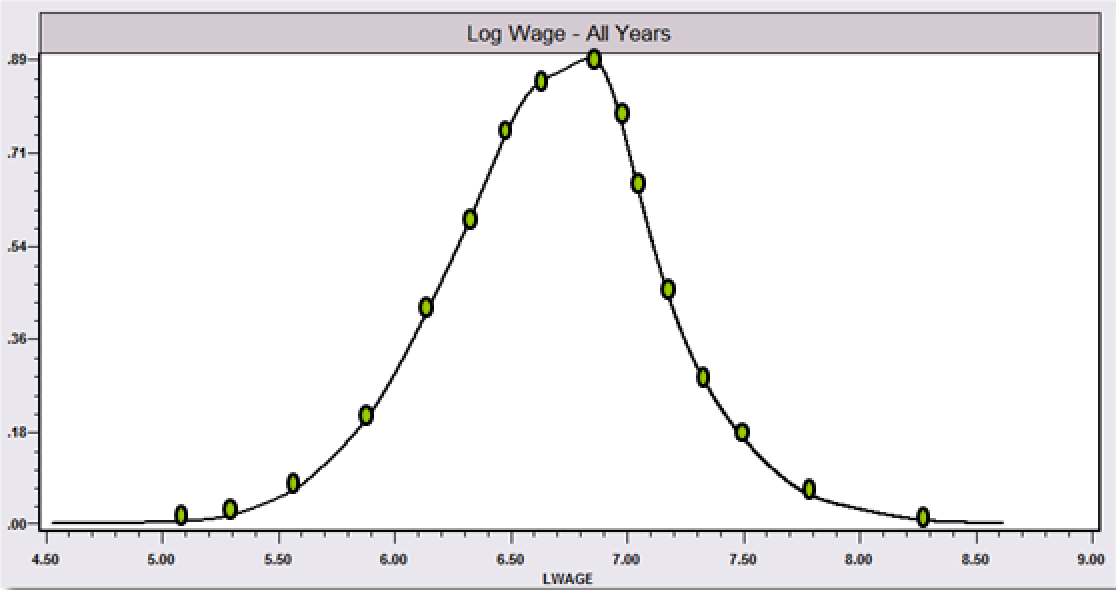
$$= \frac{1}{n}\sum_{i=1}^{n}Q(x_i \mid x_m^*, B), \text{ for a set of points } x_m^*$$

$B =$ "bandwidth"

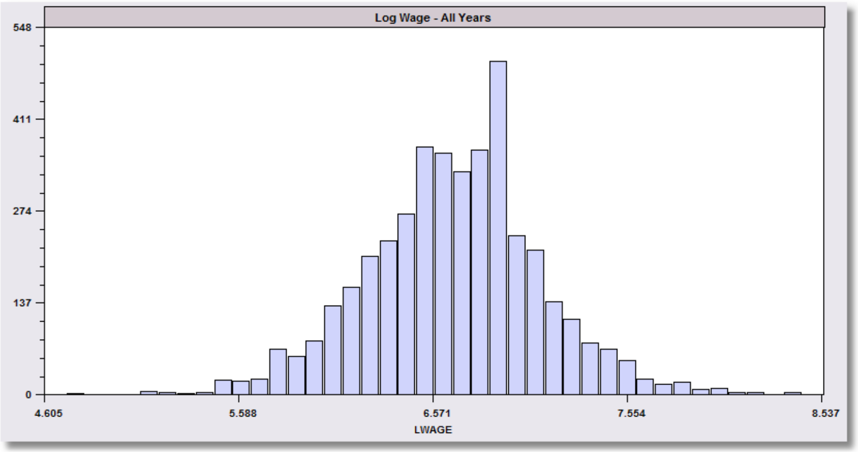
$K =$ the kernel function

$x^* =$ the point at which the density is approximated.

$\hat{f}(x^*)$ is an estimator of $f(x^*)$



The kernel density estimator is a histogram (of sorts).



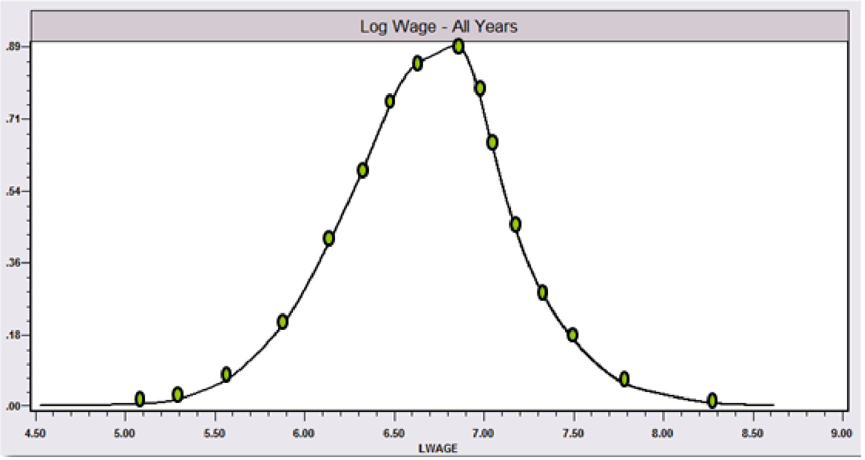
$$\hat{f}(x_m^*) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{B}K\left[\frac{x_i - x_m^*}{B}\right], \text{ for a set of points } x_m^*$$

B = "bandwidth" chosen by the analyst

K = the kernel function, such as the normal or logistic pdf (or one of several others)

x* = the point at which the density is approximated.

This is essentially a histogram with small bins.



JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **26**: 549–579 (2011)
Published online 14 December 2009 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1134

HOW DOES HETEROGENEITY SHAPE THE SOCIOECONOMIC GRADIENT IN HEALTH SATISFACTION?

ANDREW M. JONES[a] AND STEFANIE SCHURER[b]*

[a] *Department of Economics and Related Studies, University of York, York, UK*
[b] *Melbourne Institute of Applied Economic and Social Research, University of Melbourne, Parkville, Victoria, Australia*

SUMMARY

Individual heterogeneity plays a key role in explaining variation in self-reported health and its socioeconomic gradient. It is hypothesised that the influence of this heterogeneity varies over levels of health and increases over the life cycle. These hypotheses are tested by applying a threshold-specific alternative to the conditional fixed-effects logit and longitudinal data from Germany. Our results suggest that income influences health at the lower end, but not at the higher end of the health distribution, once unobservable factors are controlled for. The underlying assumptions of the statistical model matter for this conclusion, in particular for the older age groups.

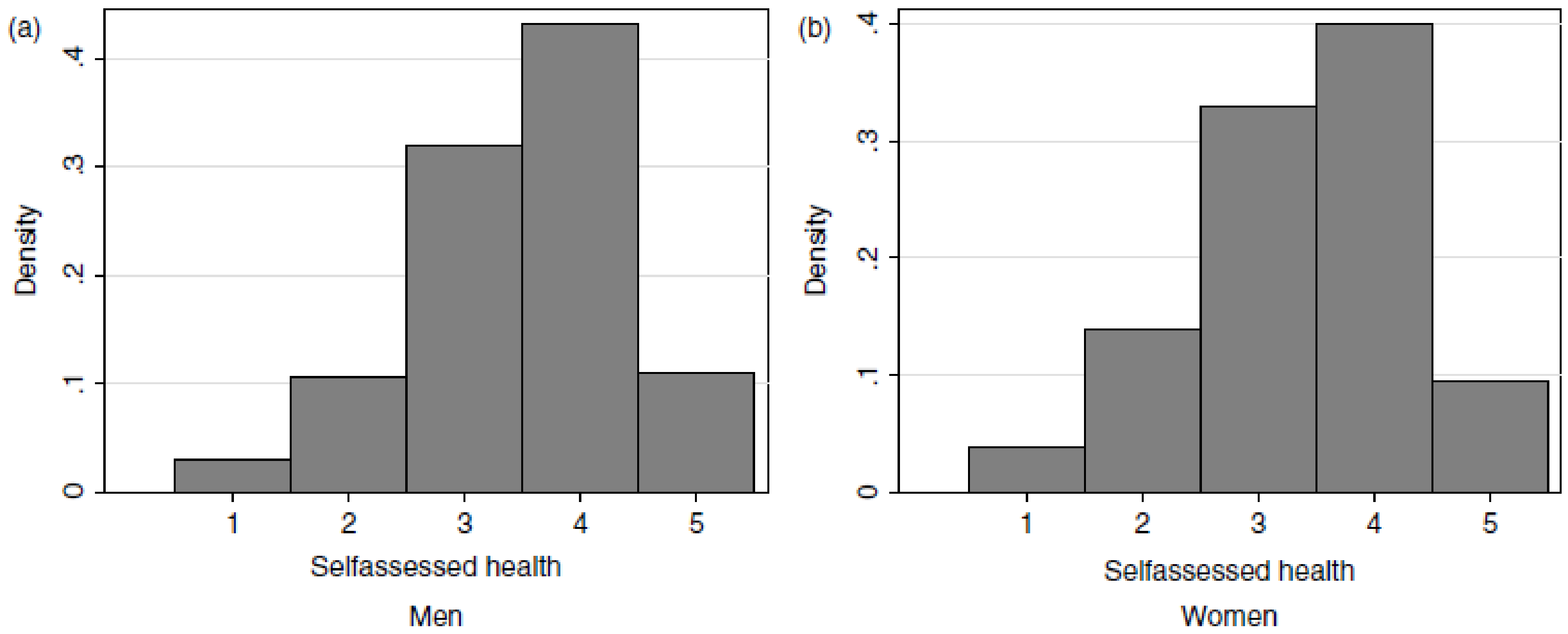
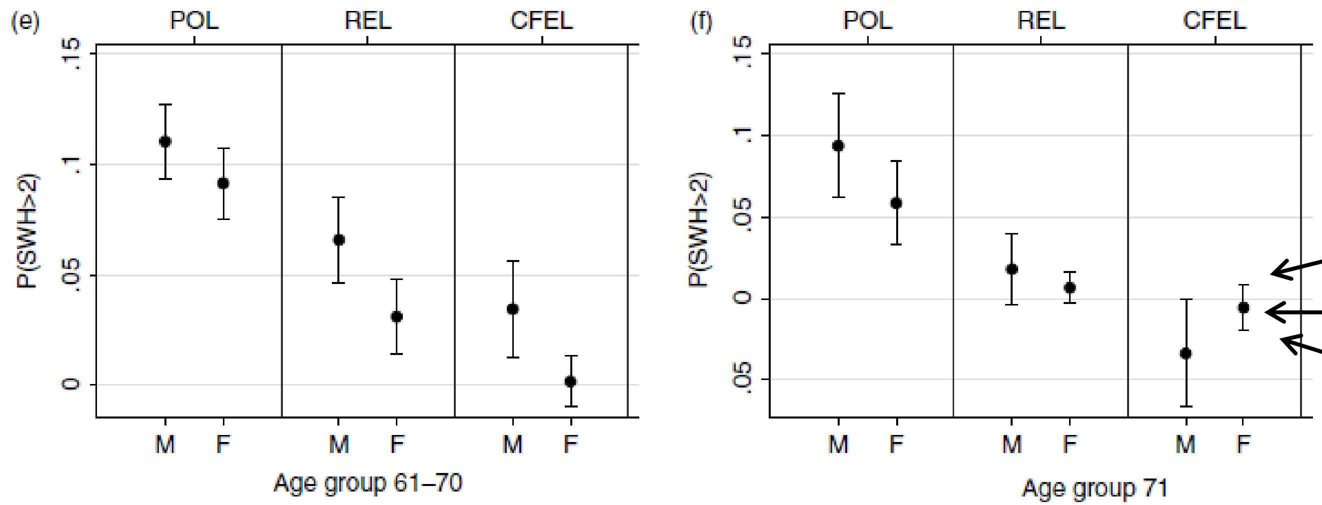


Figure 2. Distribution of health measure: self-assessed health (SAH)

From Jones and Schurer (2011) Stylized Box Plot



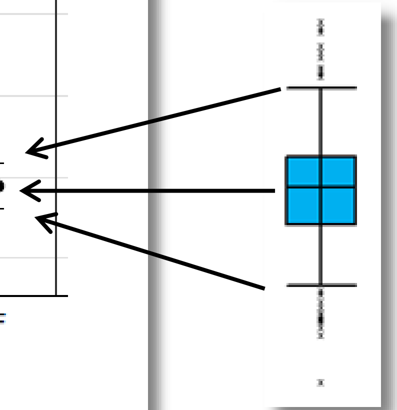


Figure 4. Marginal effects of income on the probability to report satisfaction with health greater than two for both men (M) and women (W) obtained from three different models: pooled ordered logit (POL), random-effects logit (REL), and conditional fixed-effects logit (CFEL)

From Jones and Schurer (2011)

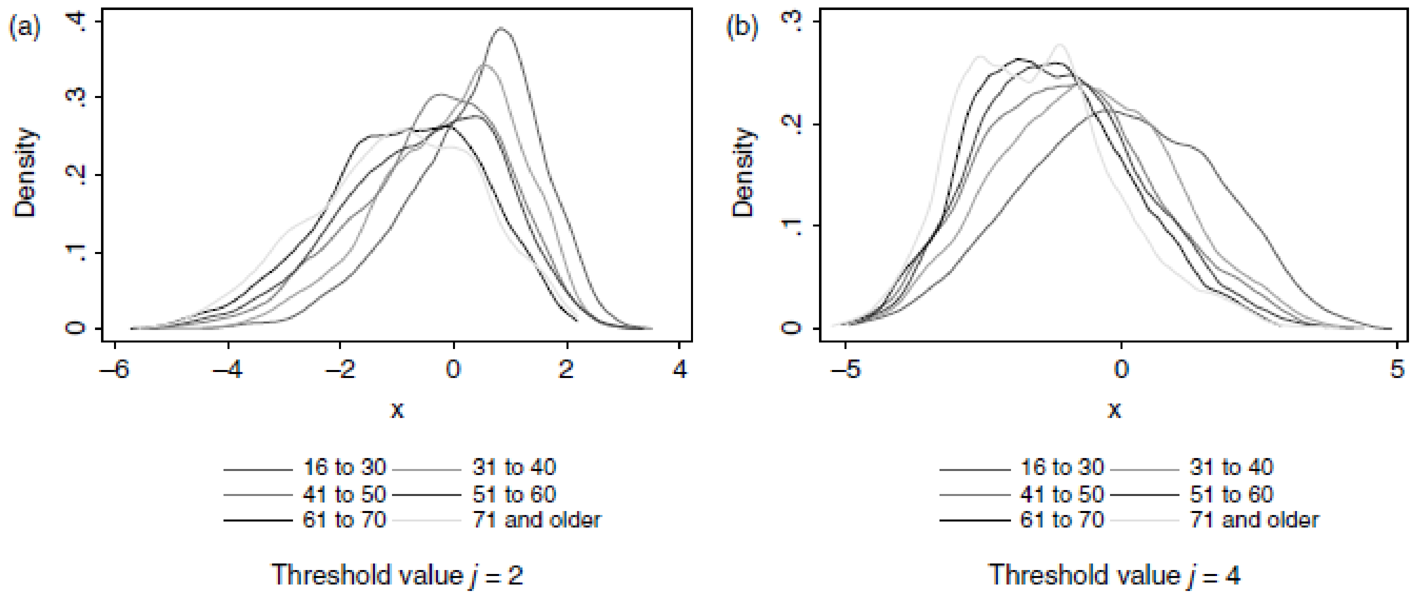
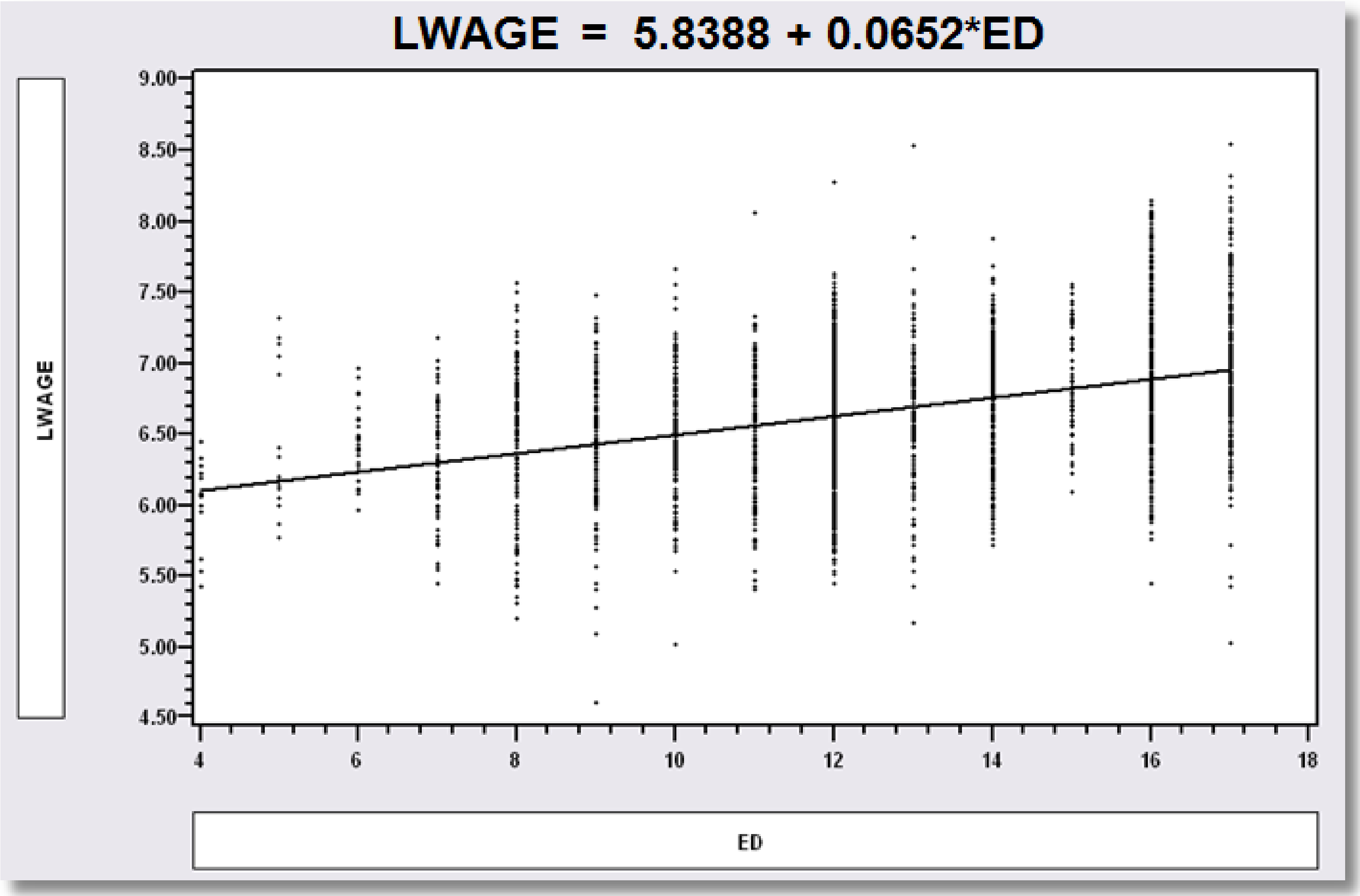


Figure 7. Probability distribution of individual fixed effect obtained from the conditional fixed-effects logit (CFEL), when the threshold values are $j = 2$ and $j = 4$ for the sample of men

Objective: Impact of Education on (log) Wage

- **Specification**: What is the right model to use to analyze this association?
- **Estimation**
- **Inference**
- **Analysis**

Simple Linear Regression



Multiple Regression

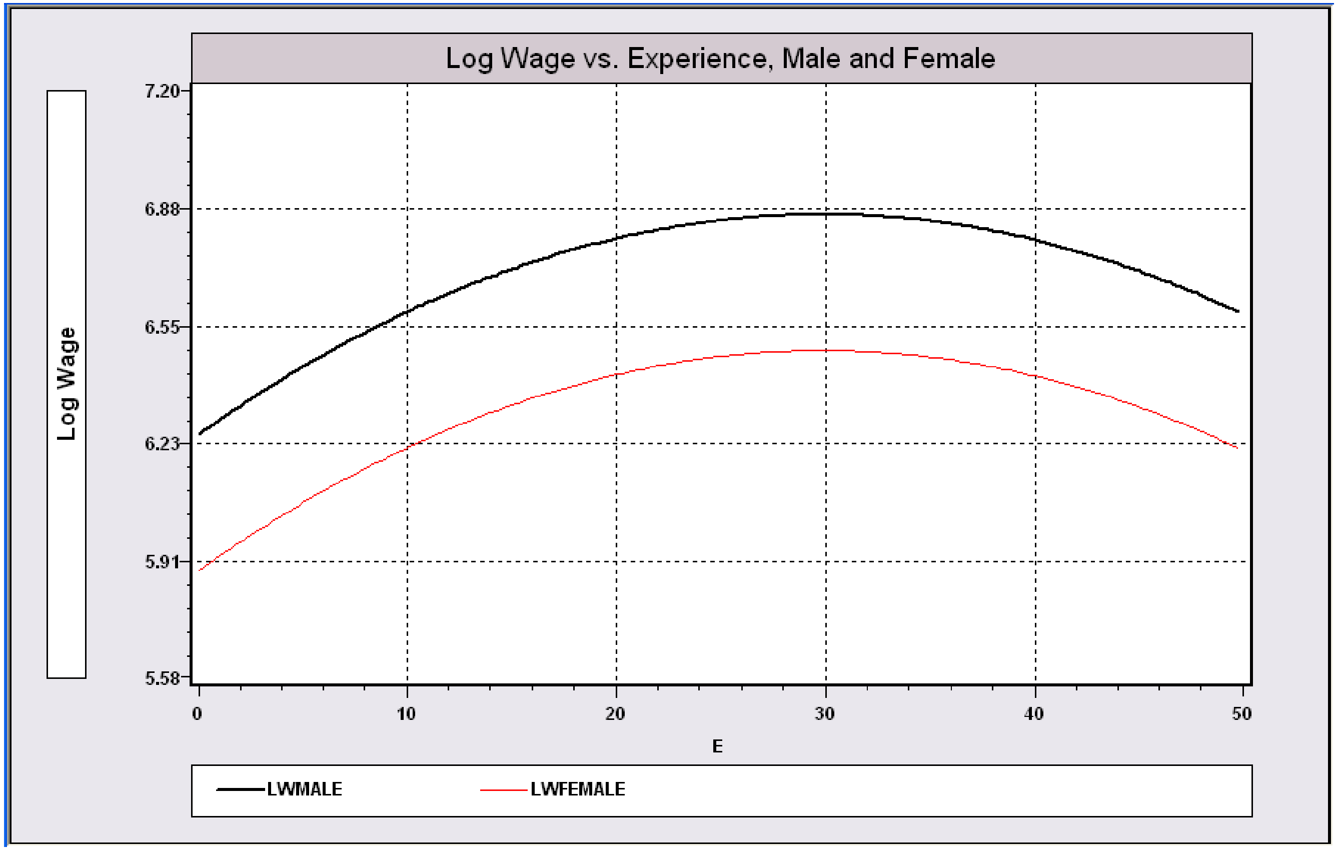
```
----------------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LWAGE    Mean                    =       6.67635
             Standard deviation      =        .46151
----------   No. of observations     =          4165  DegFreedom   Mean square
Regression   Sum of Squares          =       345.763          9      38.41812
Residual     Sum of Squares          =       541.142       4155        .13024
Total        Sum of Squares          =       886.905       4164        .21299
----------   Standard error of e     =        .36089  Root MSE          .36045
Fit          R-squared               =        .38985  R-bar squared     .38853
Model test   F[  9,  4155]           =     294.98231  Prob F > F*       .00000
--------+-------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------------
Constant|     5.44028***     .07208     75.48  .0000     5.29902    5.58155
      ED|      .05682***     .00267     21.25  .0000      .05158     .06207
     EXP|      .01040***     .00054     19.37  .0000      .00935     .01145
     WKS|      .00525***     .00111      4.71  .0000      .00306     .00743
     OCC|     -.14867***     .01507     -9.87  .0000     -.17819    -.11914
   SOUTH|     -.07024***     .01279     -5.49  .0000     -.09530    -.04517
    SMSA|      .13241***     .01235     10.72  .0000      .10820     .15663
      MS|      .08568***     .02108      4.06  .0000      .04435     .12700
     FEM|     -.37561***     .02577    -14.58  .0000     -.42611    -.32511
   UNION|      .09995***     .01318      7.58  .0000      .07411     .12579
--------+-------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

Nonlinear Specification: Quadratic Effect of Experience

```
-----------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=LWAGE     Mean                    =         6.67635
              Standard deviation      =          .46151
----------    No. of observations     =            4165  DegFreedom   Mean square
Regression    Sum of Squares          =         370.955          10      37.09546
Residual      Sum of Squares          =         515.950        4154        .12421
Total         Sum of Squares          =         886.905        4164        .21299
----------    Standard error of e     =          .35243  Root MSE          .35196
Fit           R-squared               =          .41826  R-bar squared     .41686
Model test    F[ 10,   4154]          =       298.66153  Prob F > F*       .00000
--------+--------------------------------------------------------------------
        |                   Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error        z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    5.24547***      .07170     73.15  .0000     5.10493    5.38600
      ED|     .05654***      .00261     21.64  .0000      .05142     .06166
     EXP|     .04045***      .00217     18.61  .0000      .03619     .04471
 EXP*EXP|    -.00068***   .4783D-04    -14.24  .0000     -.00077    -.00059
     WKS|     .00449***      .00109      4.12  .0000      .00235     .00662
     OCC|    -.14053***      .01472     -9.54  .0000     -.16939    -.11167
   SOUTH|    -.07210***      .01249     -5.77  .0000     -.09658    -.04762
    SMSA|     .13901***      .01207     11.51  .0000      .11534     .16267
      MS|     .06736***      .02063      3.26  .0011      .02692     .10779
     FEM|    -.38922***      .02518    -15.46  .0000     -.43857    -.33987
   UNION|     .09015***      .01289      6.99  .0000      .06488     .11542
--------+--------------------------------------------------------------------
nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Model Building in Econometrics

A Model Relating (Log)Wage to Gender and Education & Experience



Partial Effects
Coefficients may not tell the story

```
----------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=LWAGE     Mean                    =      6.67635
              Standard deviation      =       .46151
----------    No. of observations     =         4165  DegFreedom   Mean square
Regression    Sum of Squares          =      378.218          11      34.38347
Residual      Sum of Squares          =      508.687        4153        .12249
Total         Sum of Squares          =      886.905        4164        .21299
----------    Standard error of e     =       .34998  Root MSE          .34948
Fit           R-squared               =       .42645  R-bar squared     .42493
Model test    F[ 11,   4153]          =    280.71214  Prob F > F*       .00000
--------+-------------------------------------------------------------------
```

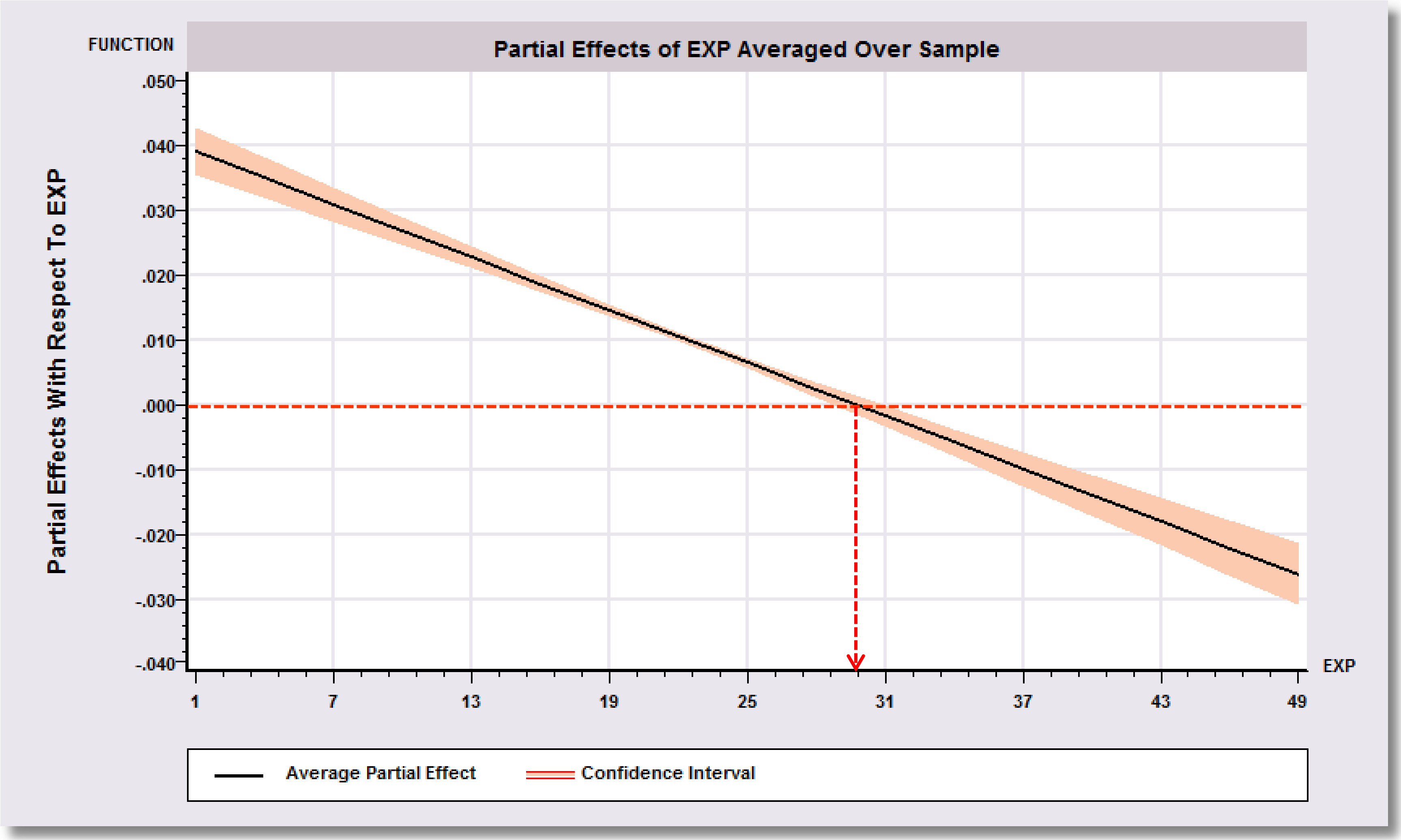
```
 Constant|      5.24547***
       ED|       .05654***
      EXP|       .04045***
  EXP*EXP|     -.00068***
      WKS|       .00449***
      OCC|     -.14053***
    SOUTH|     -.07210***
     SMSA|       .13901***
       MS|       .06736***
      FEM|     -.38922***
    UNION|       .09015***
```

Education .05654

Experience .04045 - 2*.00068*Exp

FEM -.38922

Effect of Experience = .04045 - 2*.00068*Exp
Positive from 1 to 30, negative after.



Interaction Effect
Gender Difference in Partial Effects

```
----------------------------------------------------------------------------
Ordinary      least squares regression ............
LHS=LWAGE     Mean                     =       6.67635
              Standard deviation       =        .46151
----------    No. of observations      =          4165  DegFreedom   Mean square
Regression    Sum of Squares           =       347.213          10      34.72132
Residual      Sum of Squares           =       539.692        4154        .12992
Total         Sum of Squares           =       886.905        4164        .21299
----------    Standard error of e      =        .36045  Root MSE          .35997
Fit           R-squared                =        .39149  R-bar squared     .39002
Model test    F[ 10,  4154]            =     267.24949  Prob F > F*       .00000
--------+-------------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
   LWAGE|  Coefficient       Error       z    |z|>Z*          Interval
--------+-------------------------------------------------------------------
Constant|    5.47075***     .07256    75.39  .0000     5.32853    5.61298
      ED|     .05458***     .00275    19.81  .0000      .04918     .05998
     EXP|     .01035***     .00054    19.29  .0000      .00930     .01140
     WKS|     .00528***     .00111     4.74  .0000      .00310     .00746
     OCC|    -.14659***     .01506    -9.73  .0000     -.17611    -.11707
   SOUTH|    -.07176***     .01278    -5.61  .0000     -.09682    -.04671
    SMSA|     .13351***     .01234    10.82  .0000      .10932     .15770
      MS|     .08392***     .02107     3.98  .0001      .04263     .12520
     FEM|    -.67961***     .09456    -7.19  .0000     -.86495    -.49427
   UNION|     .09496***     .01325     7.17  .0000      .06899     .12093
  ED*FEM|     .02350***     .00703     3.34  .0008      .00971     .03729
--------+-------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------------
```

Partial Effect of a Year of Education

$$\partial E[\log Wage]/\partial ED = \beta_{ED} + \beta_{ED*FEM} * FEM$$

$$= 0.05458 + 0.02350*FEM$$

Note, the effect is positive.

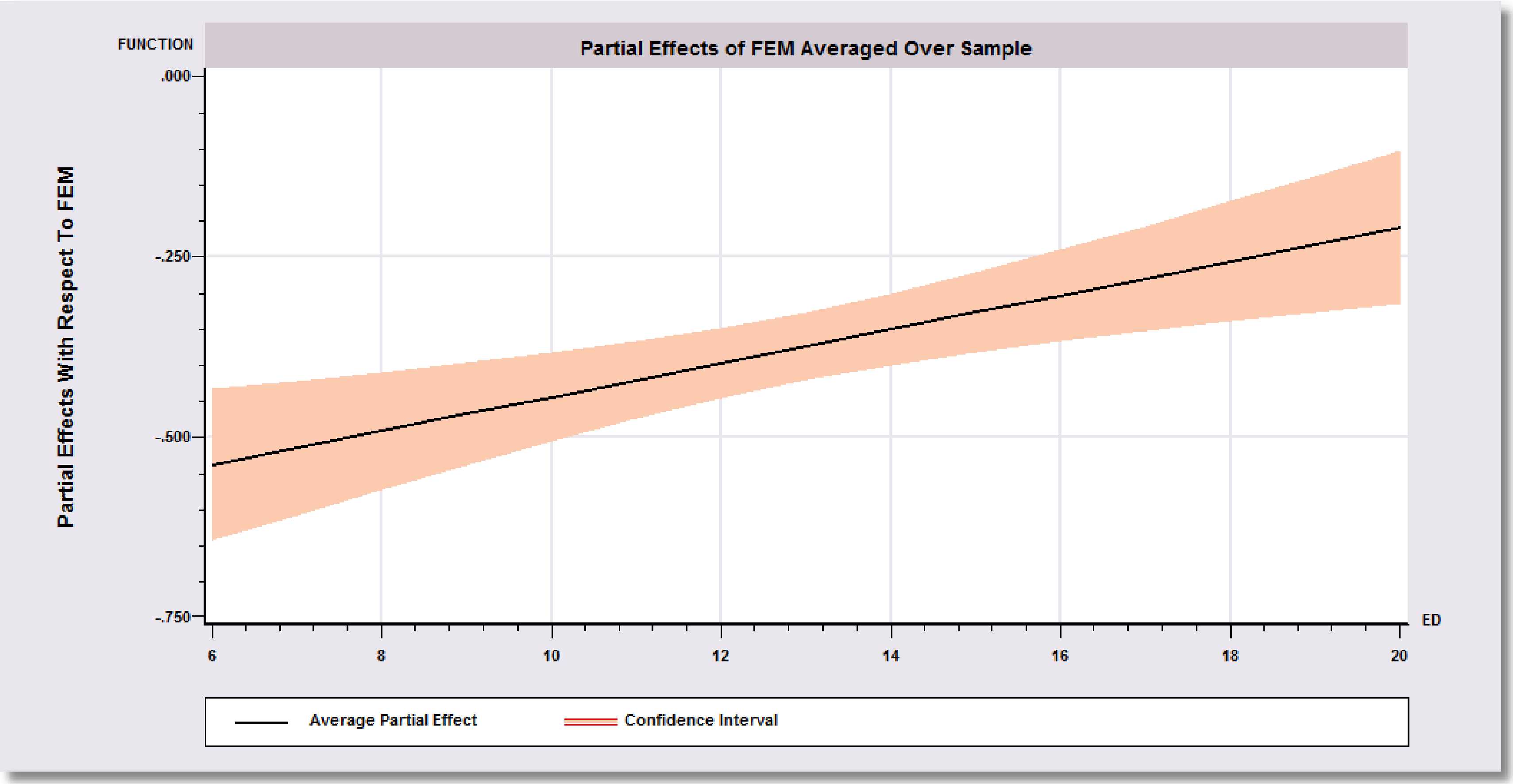
The effect 43% is larger for women.

```
--------------------------------------------------------------------------
Partial Effects  Analysis for Linear Regression Function
--------------------------------------------------------------------------
Effects on function with respect to ED
Results are computed by average over sample observations
Partial effects for continuous ED       computed by differentiation
Effect is computed as derivative     = df(.)/dx
--------------------------------------------------------------------------
df/dED                 Partial     Standard
(Delta method)         Effect       Error      |t|   95% Confidence Interval
--------------------------------------------------------------------------
APE. Function           .05723      .00267   21.40      .05199       .06247
--------------------------------------------------------------------------
FEM      =    .00 ------------------------------------------------------
Average effect          .05458      .00275   19.81      .04918       .05998
--------------------------------------------------------------------------
FEM      =   1.00 ------------------------------------------------------
Average effect          .07808      .00690   11.32      .06456       .09161
```

Gender Effect Varies by Years of Education

$$\Delta E[\text{logWage}]/\nabla \text{FEM} = \beta_{FEM} + \beta_{ED*FEM} *ED$$

$$= -0.67961 + 0.0235*ED$$



Analysts are interested in interaction effects in models.