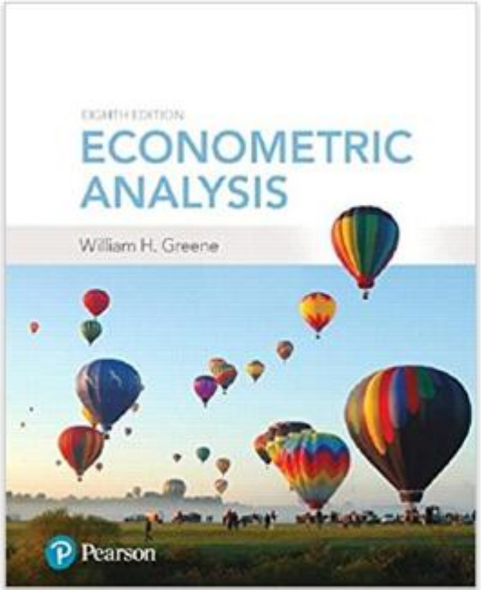
Econometrics I

Professor William Greene
Stern School of Business
Department of Economics



Econometrics I

Part 10 – Interval Estimation and Prediction

Interval Estimation

- **b** = point estimator of β
- We acknowledge the sampling variability.
  - Estimated sampling variance

logG	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
Constant	-23.9398*	12.19453	-1.96	.0593	-48.8804	1.0008
YEAR	.00902	.00677	1.33	.1931	-.00483	.02287
logPG	-.47878***	.06162	-7.77	.0000	-.60482	-.35275
logY	1.27365***	.15010	8.49	.0000	.96666	1.58064
logPN	1.13036***	.18906	5.98	.0000	.74369	1.51703
logPD	.79402***	.22581	3.52	.0015	.33217	1.25586
logPS	-1.19347***	.20592	-5.80	.0000	-1.61462	-.77233

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

- Point estimate is only the best single guess
- Form an interval, or range of plausible values
- Plausible ⇔ likely values with acceptable degree of probability.
- To assign probabilities, we require a distribution for the variation of the estimator.
- The role of the normality assumption for ε

Robust Inference: Confidence Interval for β_k

b_k = the point estimate

Std.Err[b_k] = sqr{Estimated Asymptotic Variance of b_k} = v_k

The matrix may be any robust or appropriate covariance matrix.

- $b_k \sim$ Asy.N[β_k, v_k^2]
- $(b_k-\beta_k)/v_k \sim$ Asy.N[0,1]

Consider a range of plausible values of β_k given the point estimate b_k. $b_k \pm$ sampling error.

- Measured in standard error units,
- $|(b_k - \beta_k)/ v_k| < z^*$
- Larger z* ➔ greater probability ("confidence")
- Given normality, e.g., z* = 1.96 ➔ 95%, z*=1.645➔90%
- Plausible range for β_k then is $b_k \pm z^* v_k$

Critical Values for the Confidence Interval

Assume normality of $\boldsymbol{\varepsilon}$:

- $b_k \sim N[\beta_k, v_k^2]$ for the true β_k.
- $(b_k - \beta_k)/v_k \sim N[0,1]$

$v_k = [s^2(\mathbf{X'X})^{-1}]_{kk}$ $(b_k - \beta_k)/\text{Est.}(v_k) \sim t[n-K]$.

Use critical values from t[n-K] distribution instead of standard normal. Will be the same as normal if $n \geq 100$.

Based on asymptotic results or any robust covariance matrix estimator, use N[0,1].

Confidence Interval

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=logG    Mean                 =          5.39299
            Standard deviation   =           .24878
----------  No. of observations  =               36   DegFreedom   Mean square
Regression  Sum of Squares       =          2.15567            6        .35928
Residual    Sum of Squares       =       .105227E-01          29        .00036
Total       Sum of Squares       =          2.16619           35        .06189
----------  Standard error of e  =           .01905   Root MSE           .01710
Fit         R-squared            =           .99514   R-bar squared      .99414
Model test  F[  6,    29]        =        990.14814   Prob F > F*        .00000
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
    logG|  Coefficient       Error        t   |t|>T*         Interval
--------+--------------------------------------------------------------------
Constant|   -23.9398*      12.19453    -1.96  .0593    -48.8804     1.0008
    YEAR|     .00902         .00677     1.33  .1931     -.00483     .02287
   logPG|    -.47878***      .06162    -7.77  .0000     -.60482    -.35275
    logY|    1.27365***      .15010     8.49  .0000      .96666    1.58064
   logPN|    1.13036***      .18906     5.98  .0000      .74369    1.51703
   logPD|     .79402***      .22581     3.52  .0015      .33217    1.25586
   logPS|   -1.19347***      .20592    -5.80  .0000    -1.61462    -.77233
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Critical t[.975,29] = 2.045

Confidence interval based on t: $1.27365 \pm 2.045 * .1501$

Confidence interval based on normal: $1.27365 \pm 1.960 * .1501$

Specification and Functional Form: Interaction Effects

Population

$$y = \beta_1 + \beta_2 x + \beta_3 z + \beta_4 xz + \varepsilon$$

$$\delta_x = \frac{\partial E[y \mid x, z]}{\partial x} = \beta_2 + \beta_4 z$$

Estimators

$$\hat{y} = b_1 + b_2 x + b_3 z + b_4 xz$$

$$\hat{\delta}_x = b_2 + b_4 z$$

Estimator of the variance of $\hat{\delta}_x$

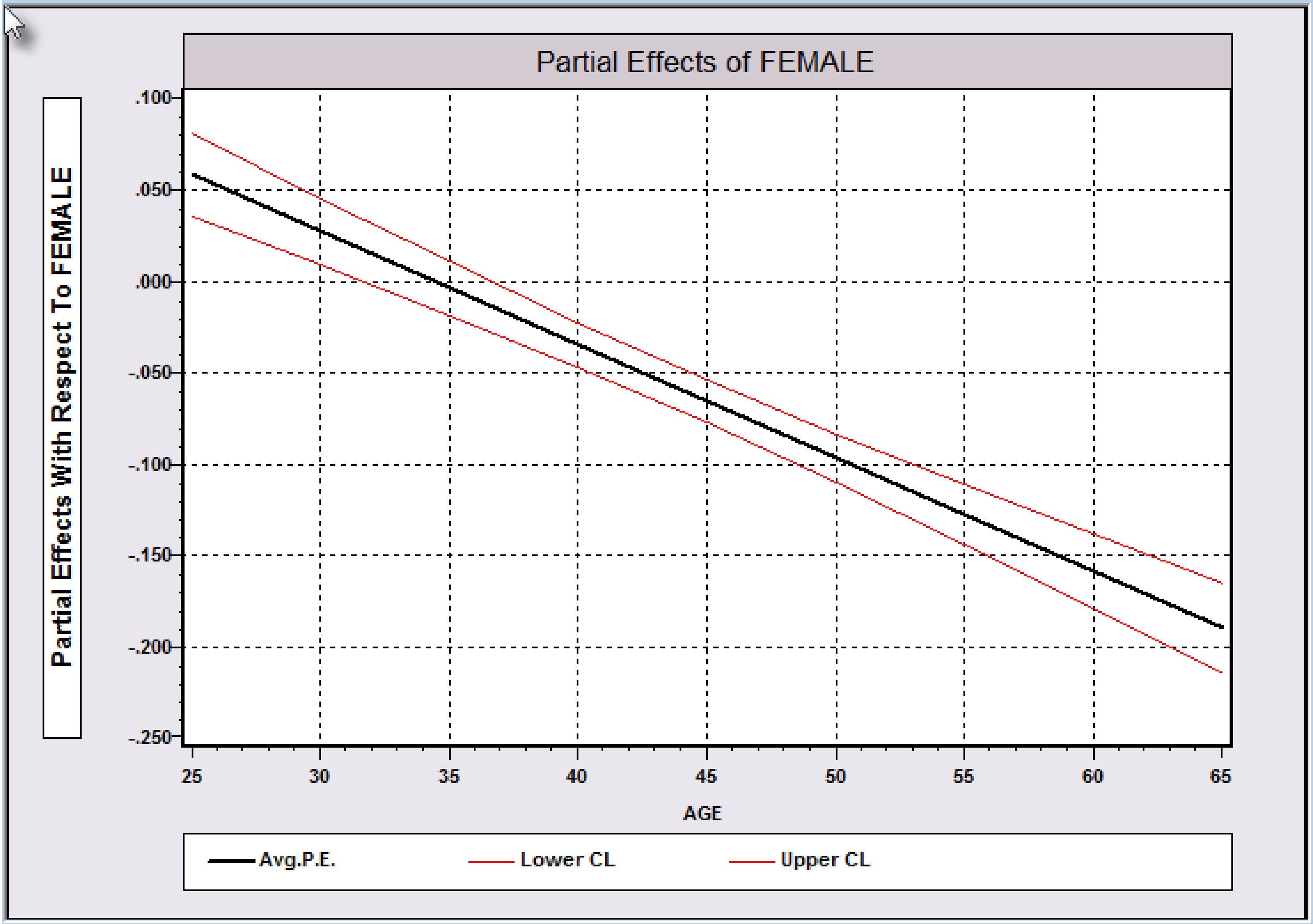
$$Est.Var[\hat{\delta}_x] = Var[b_2] + z^2 Var[b_4] + 2zCov[b_2, b_4]$$

Interaction Effect

```
-----------------------------------------------------------------------
Ordinary     least squares regression ............
LHS=LOGY     Mean                     =     -1.15746
             Standard deviation       =       .49149
             Number of observs.       =        27322
Model size   Parameters               =            4
             Degrees of freedom       =        27318
Residuals    Sum of squares           =   6540.45988
             Standard error of e      =       .48931
Fit          R-squared                =       .00896
             Adjusted R-squared       =       .00885
Model test   F[  3, 27318] (prob) =    82.4(.0000)
--------+--------------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+--------------------------------------------------------------
Constant|   -1.22592***       .01605      -76.376   .0000
     AGE|     .00227***       .00036        6.240   .0000      43.5272
  FEMALE|     .21239***       .02363        8.987   .0000       .47881
 AGE_FEM|    -.00620***       .00052      -11.819   .0000      21.2960
--------+--------------------------------------------------------------
```

Do women earn more than men (in this sample?) The +.21239 coefficient on FEMALE would suggest so. But, the female "difference" is +.21239 - .00620*Age. At average Age, the effect is .21239 - .00620(43.5272) = -.05748.

```
--------------------------------------------------------------------
Partial Effects  Analysis for Linear Regression Function
--------------------------------------------------------------------
Effects on function with respect to FEMALE
Results are computed by average over sample observations
Partial effects for binary var FEMALE   computed by first difference
--------------------------------------------------------------------
df/dFEMALE            Partial    Standard         95% Confidence Interval
(Delta method)        Effect      Error     |t|
--------------------------------------------------------------------
APE. Function         -.05750     .00595   9.67     -.06916     -.04585
AGE     = 25.00        .05737     .01143   5.02      .03498      .07977
AGE     = 30.00        .02637     .00929   2.84      .00817      .04457
AGE     = 35.00       -.00463     .00746    .62     -.01926      .01000
AGE     = 40.00       -.03563     .00624   5.71     -.04786     -.02341
AGE     = 45.00       -.06664     .00599  11.12     -.07838     -.05490
AGE     = 50.00       -.09764 |   .00683  14.30     -.11102     -.08426
AGE     = 55.00       -.12864     .00843  15.26     -.14517     -.11211
AGE     = 60.00       -.15964     .01046  15.27     -.18014     -.13915
AGE     = 65.00       -.19065     .01270  15.01     -.21555     -.16575
```


Partial Effects of FEMALE
Partial Effects With Respect To FEMALE
.100
.050
.000
-.050
-.100
-.150
-.200
-.250
25
30
35
40
45
50
55
60
65
AGE
Avg.P.E.
Lower CL
Upper CL

Bootstrap Confidence Interval For a Coefficient

```
regr;lhs=logg;rhs=logi,one,logpg,logpnc,logpuc,logppt$
calc;bi = b(1) ; bbar=0 $
matrix ; bir=init(1000,1,0.0)$
proc$
draw    ; n=52 ; replacement $
regress ; quietly
        ; lhs=logg;rhs=logi,one,logpg,logpnc,logpuc,logppt$
matrix  ; bir(i)=b(1) $
calc    ; bbar=bbar + 1/1000 *b(1) $
endproc $
exec;i=1,1000$
matrix ; ones = init(1000,1,1.0)
       ; brep = bi*ones + bir-bbar*ones $
quantiles ; rhs = brep $
```

Bootstrap CI for Least Squares

Ordinary	least squares regression				
LHS=LOGG	Mean	=	-9.94245		
	Standard deviation	=	.23881		
----------	No. of observations	=	52	DegFreedom	Mean square
Regression	Sum of Squares	=	2.79360	5	.55872
Residual	Sum of Squares	=	.114979	46	.00250
Total	Sum of Squares	=	2.90858	51	.05703
----------	Standard error of e	=	.05000	Root MSE	.04702
Fit	R-squared	=	.96047	R-bar squared	.95617
Model test	F[5, 46]	=	223.52968	Prob F > F*	.00000

LOGG	Coefficient	Standard Error	t	Prob. |t|>T*	95% Confidence Interval	
LOGI	1.28609***	.14566	8.83	.0000	.99289	1.57930
Constant	-20.9548***	1.52568	-13.73	.0000	-24.0259	-17.8838
LOGPG	-.02797	.04338	-.64	.5222	-.11529	.05934
LOGPNC	-.15576	.21002	-.74	.4621	-.57852	.26699
LOGPUC	.02850	.10199	.28	.7811	-.17680	.23381
LOGPPT	-.18283	.11908	-1.54	.1315	-.42252	.05686

Percentiles	BREP:1
Sample size	1000
Min.	.77187
01th	.94799
*025	1.01719
05th	1.06692
10th	1.12893
20th	1.19682
25th	1.21584
30th	1.23299
40th	1.25748
Med.	1.28525
60th	1.31333
70th	1.34672
75th	1.36097
80th	1.37399
90th	1.43407
95th	1.49031
*975	1.55168
99th	1.64709
Max.	1.96372

Bootstrap CI for Least Absolute Deviations

```
--------------------------------------------------------------------------------
Least absolute deviations estimator...............
Nonlinear     least squares regression ............
LHS=LOGG      Mean                     =       -9.95023
              Standard deviation       =         .24202
              Number of observs.       =             52
Model size    Parameters               =              6
              Degrees of freedom       =             46
Residuals     Sum of squares           =        .119690
              Standard error of e      =         .05101
Fit           R-squared                =         .95993
              Adjusted R-squared       =         .95558
Model test    F[  5,    46] (prob) =   220.4(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Sum of absolute deviations         =      1.6665854
--------+-----------------------------------------------------------------------
        |                Standard            Prob.       95% Confidence
    LOGG|  Coefficient     Error       z    |z|>Z*          Interval
--------+-----------------------------------------------------------------------
        |Covariance matrix based on   50 replications.
    LOGI|     1.20636***     .16869     7.15  .0000        .87573    1.53698
Constant|   -19.6510***     1.75404   -11.20  .0000     -23.0889   -16.2132
   LOGPG|      .04376        .03901     1.12  .2619      -.03269     .12022
  LOGPNC|     -.50766        .34833    -1.46  .1450     -1.19038     .17505
  LOGPUC|      .04250        .17786      .24  .8111      -.30610     .39110
  LOGPPT|     -.01514        .16300     -.09  .9260      -.33461     .30433
--------+-----------------------------------------------------------------------
```

```
------------------------------
Percentiles          BREP:1
Sample size           1000
------------------------------
Min.                 .11903
01th                 .53445
*025                 .72084
05th                 .83721
10th                 .93752
20th                1.04461
25th                1.08954
30th                1.11451
40th                1.16544
Med.                1.22374
60th                1.26602
70th                1.31469
75th                1.34340
80th                1.37143
90th                1.45886
95th                1.54635
*975                1.60223
99th                1.70422
Max.                2.16943
------------------------------
```

Bootstrapped Confidence Intervals

Estimate Norm(β)=$(\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2)^{1/2}$

Untitled 1 *

Insert Name:

```
?**********************************************
? Delta Method
?**********************************************
Namelist ; x=one,x1,x2,x3,x4$
Regress  ; lhs=yit;rhs=x$
Wald     ; parameters = b
         ; covariance = varb
         ; labels=a,b1,b2,b3,b4
         ; fn1 = sqr(b1^2+b2^2+b3^2+b4^2) $
?**********************************************
? Bootstrap Method
?**********************************************
Proc$
Regress  ; quietly ; lhs=yit ; rhs=x$
Matrix   ; beta=b(2:5);bnorm=sqrt(beta'beta)$
Endproc$
Execute  ; n=500 ; bootstrap = bnorm$
Create   ; bootnorm=bootstrp$
Dstat    ; Rhs=bootnorm ; quantiles$
```

```
--------+--------------------------------------------------------------------
        |                   Standard                                 Missing
Variable|        Mean       Deviation      Minimum      Maximum   Cases  Values
--------+--------------------------------------------------------------------
BOOTNORM|     .748701       .012426       .715269      .791033     500     982
--------+--------------------------------------------------------------------

Descriptive Statistics for    1 variables
DSTAT results are matrix LASTDSTA in current project.
--------------------------
Percentiles      BOOTNORM
Sample size           500
--------------------------
Min.              .715269
01th              .722066
*025              .725308
05th              .727917
10th              .732863
20th              .738778
25th              .740387
30th              .742011
40th              .745226
Med.              .748275
60th              .751571
70th              .754829
75th              .757144
80th               .75874
90th              .765108
95th              .770278
*975               .77354
99th              .777276
Max.              .791033
--------------------------
```

```
-----------------------------------------------------------------------------
WALD procedure. Estimates and standard errors for nonlinear functions and
joint test of nonlinear restrictions.
Wald Statistic               =    4841.03525
Prob. from Chi-squared[ 1] =        .00000
Functions are computed at means of variables
--------+--------------------------------------------------------------------
        |                   Standard              Prob.       95% Confidence
WaldFcns|    Function        Error          z    |z|>Z*          Interval
--------+--------------------------------------------------------------------
 Fncn(1)|     .74792***     .01075      69.58   .0000       .72685     .76899
--------+--------------------------------------------------------------------
```

HEALTH ECONOMICS
Health Econ. **19**: 1063–1074 (2010)
Published online in Wiley InterScience (www.interscience.wiley.com).10.1002/hec.1647

QUANTILE REGRESSION ANALYSIS OF THE RATIONAL ADDICTION MODEL: INVESTIGATING HETEROGENEITY IN FORWARD-LOOKING BEHAVIOR

AUDREY LAPORTE[a,*], ALFIA KARIMOVA[b] and BRIAN FERGUSON[c]

[a]*Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ont., Canada*
[b]*Department of Economics, University of Toronto, Toronto, Ont., Canada*
[c]*Department of Economics, University of Guelph, Guelph, Ont., Canada*

SUMMARY

The time path of consumption from a rational addiction (RA) model contains information about an individual's tendency to be forward looking. In this paper, we use quantile regression (QR) techniques to investigate whether the tendency to be forward looking varies systematically with the level of consumption of cigarettes. Using panel data,

Coefficient on MALE dummy variable in quantile regressions

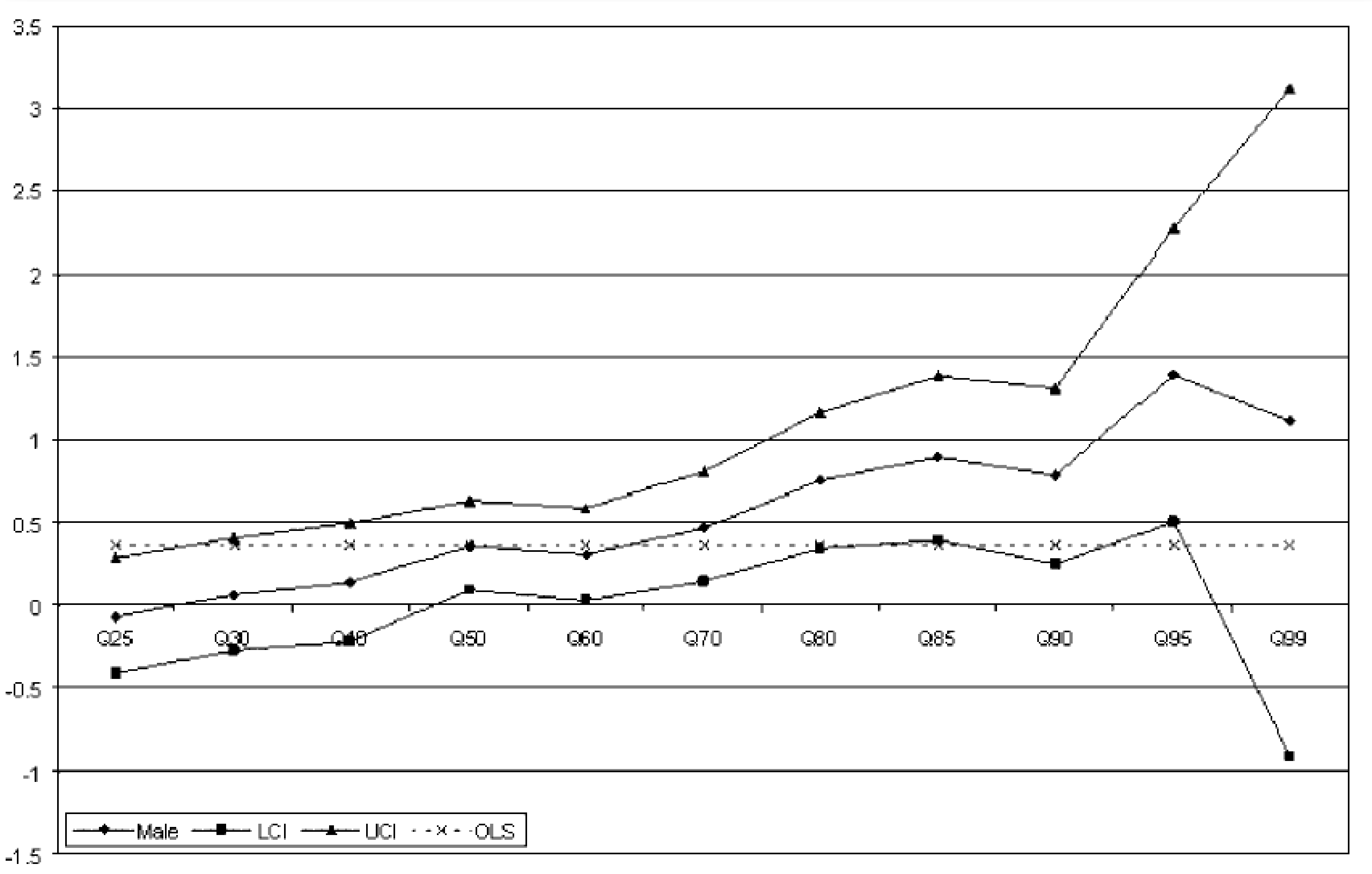


Figure 4. Male

Forecasting

- **Objective: Forecast**
- Distinction: Ex post vs. Ex ante forecasting
  - Ex post: RHS data are observed
  - Ex ante: RHS data must be forecasted
- Prediction vs. model validation.
  - Within sample prediction
  - "Hold out sample"

Prediction Intervals

Given $\mathbf{x}^0$ predict y^0.

Two cases:

Estimate $E[y|\mathbf{x}^0] = \boldsymbol{\beta}'\mathbf{x}^0$;

Predict $y^0 = \boldsymbol{\beta}'\mathbf{x}^0 + \varepsilon^0$

Obvious predictor, **b'x**0 + estimate of ε^0. Forecast ε^0 as 0, but allow for variance.

Alternative: When we predict y^0 with $\mathbf{b}'\mathbf{x}^0$, what is the 'forecast error?'

Est.y^0 - $y^0 = \mathbf{b}'\mathbf{x}^0 - \boldsymbol{\beta}'\mathbf{x}^0 - \varepsilon^0$, so the variance of the forecast error is

$$\mathbf{x}^{0\prime}\text{Var}[\mathbf{b} - \boldsymbol{\beta}]\mathbf{x}^0 + \sigma^2$$

How do we estimate this? Form a confidence interval. Two cases:

If $\mathbf{x}^0$ is a vector of constants, the variance is just $\mathbf{x}^{0\prime}$ Var[**b**] $\mathbf{x}^0$. Form confidence interval as usual.

If $\mathbf{x}^0$ had to be estimated, then we use a random variable. What is the variance of the product? (Ouch!) One possibility: Use bootstrapping.

Forecast Variance

Variance of the forecast error is

$\sigma^2 + \mathbf{x}^{0\prime}\,\text{Var}[\mathbf{b}]\mathbf{x}^0 = \sigma^2 + \sigma^2[\mathbf{x}^{0\prime}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^0]$

If the model contains a constant term, this is

$$\text{Var}[e^0] = \sigma^2\left[1 + \frac{1}{n} + \sum_{j=1}^{K-1}\sum_{k=1}^{K-1}(x_j^0 - \bar{x}_j)(x_k^0 - \bar{x}_k)(\mathbf{Z}'\mathbf{M}^0\mathbf{Z})^{jk}\right]$$

In terms squares and cross products of deviations from means. Interpretation: Forecast variance is smallest in the middle of our "experience" and increases as we move outside it.

Butterfly Effect

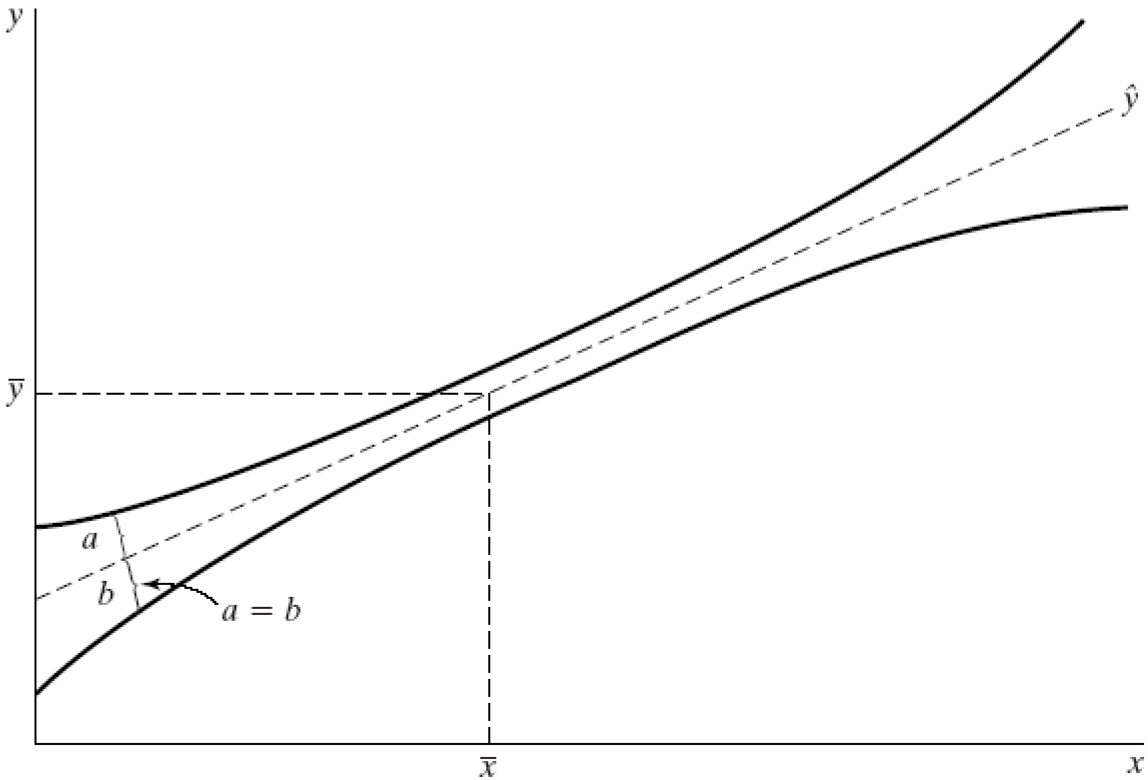
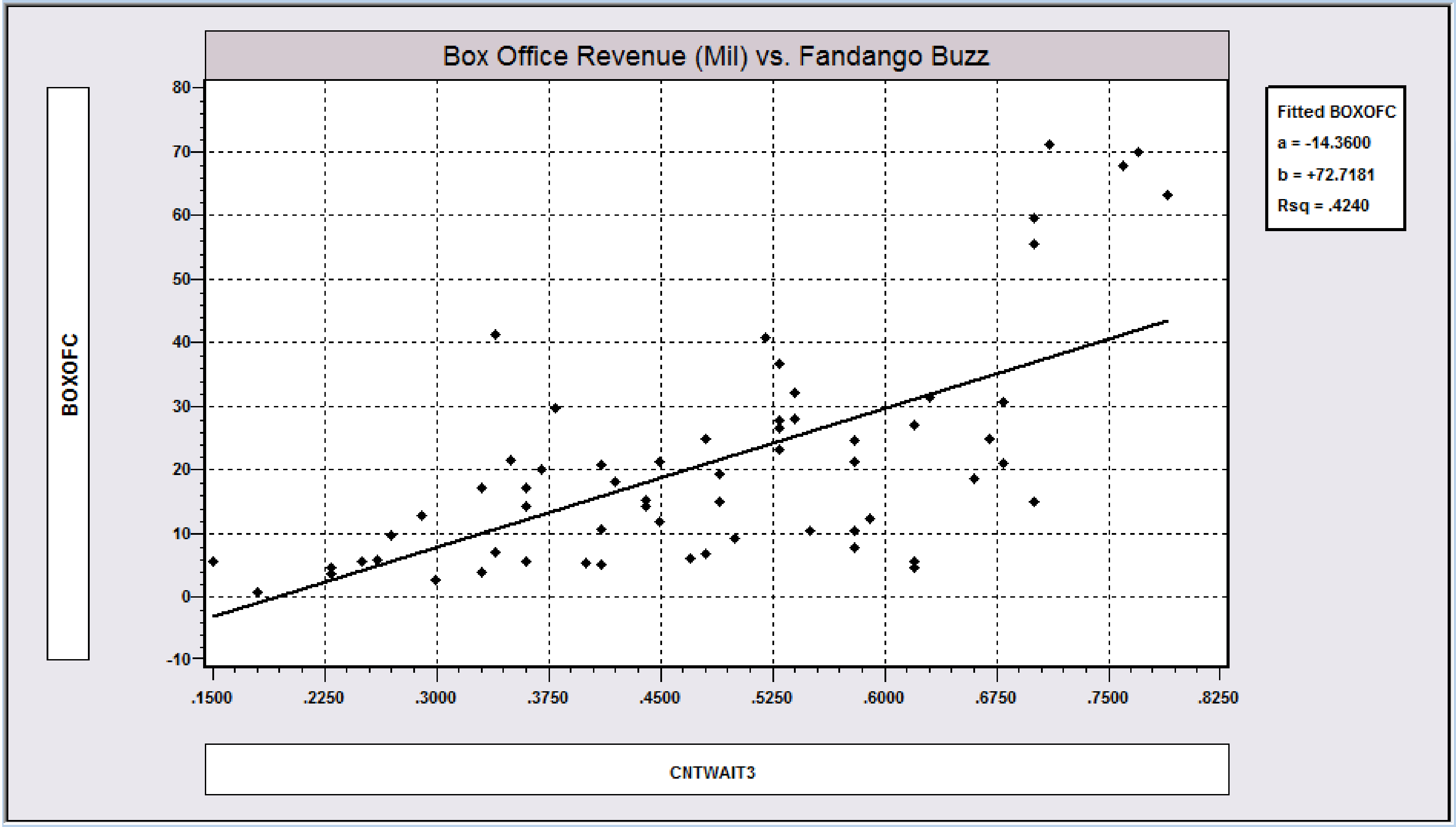


FIGURE 6.1 Prediction Intervals.

Internet Buzz Data



A Prediction Interval

Prediction includes a range of uncertainty

Point estimate: $\hat{y} = a + bx^*$

The range of uncertainty around the prediction:

$$a + bx^* \pm 1.96\sqrt{S_e^2\left(1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

The usual 95% **Due to ε** **Due to estimating α and β with a and b**

Slightly Simpler Formula for Prediction

Prediction includes a range of uncertainty

Point estimate: $\hat{y} = a + bx^*$

The range of uncertainty around the prediction:

$$a + bx^* \ \pm \ 1.96\sqrt{S_e^2\left(1+\frac{1}{n}\right) + (x^* - \bar{x})^2\left(SE(b)\right)^2}$$

Prediction from Internet Buzz Regression

Session

Regression Analysis: BoxOffice versus Buzz

```
The regression equation is
BoxOffice = - 14.4 + 72.7 Buzz


Predictor      Coef  SE Coef      T      P
Constant    -14.360    5.546  -2.59  0.012
Buzz          72.72    10.94   6.65  0.000


S = 13.3863   R-Sq = 42.4%   R-Sq(adj) = 41.4%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   7913.6  7913.6  44.16  0.000
Residual Error  60  10751.5   179.2
Total           61  18665.1

MTB >
```

$\overline{Buzz}$ = 0.48242

Max(Buzz)= 0.79

Worksheet 1 ***

↓	C1 BoxOffice	C2 Buzz	C3 SampleSplit	C4 Residual
1	19.1671	0.49	0	-2.1048
2	63.1066	0.79	1	20.0193
3	5.4016	0.36	0	-6.4169
4	67.5289	0.76	1	26.6231
5	26.2231	0.53	1	2.0425
6	69.6377	0.77	0	28.0047
7	14.8007	0.49	0	-6.4712
8	31.0698	0.63	1	-0.3826
9	12.0635	0.59	1	-16.4802
10	4.2714	0.62	1	-26.4538
11	3.4456	0.23	0	1.0804
12	0.5119	0.18	0	1.7826
13	30.4802	0.68	0	-4.6081
14	3.5317	0.33	1	-6.1053
15	14.8511	0.70	1	-21.6916

Prediction Interval for Buzz = .8

Predict Box Office for Buzz = .8

a+bx = -14.36 + 72.72(.8) = 43.82

$$\sqrt{s_e^2\left(1+\frac{1}{N}\right)+(.8-\overline{\text{Buzz}})^2 SE(b)^2}$$

$$=\sqrt{13.3863^2\left(1+\frac{1}{62}\right)+(.8-.48242)^2 10.94^2}$$

$$= 13.93$$

Interval = 43.82 $\pm$ 1.96(13.93)

= 16.52 to 71.12

Semi- and Nonparametric Estimation

Application: Stochastic Frontier Model

Production Function Regression: logY = b'x + v - u

where u is "inefficiency." u > 0. v is normally distributed.

Save for the constant term, the model is consistently estimated by OLS.

If the theory is right, the OLS residuals will be skewed to the left, rather than symmetrically distributed if they were normally distributed.

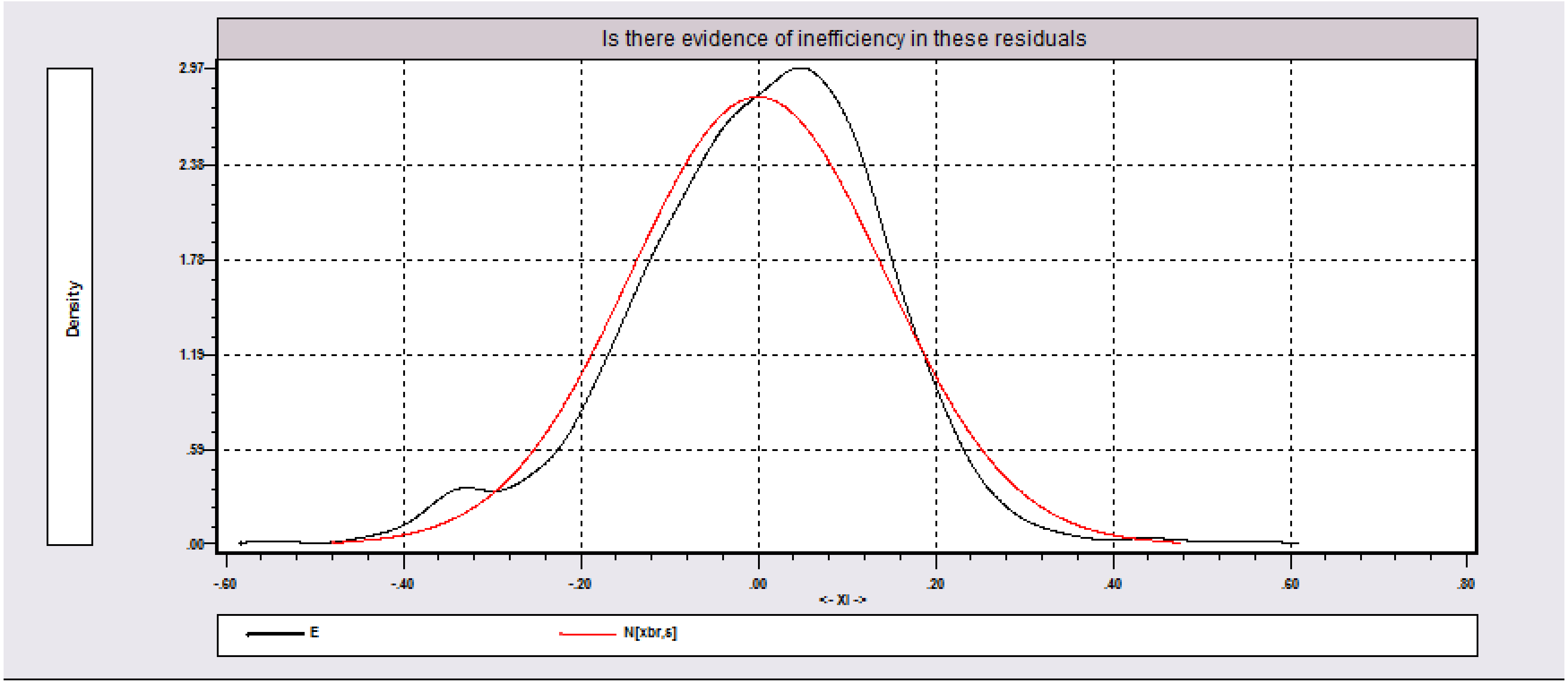
Application: Spanish dairy data used in Assignment 2

yit = log of milk production
x1 = log cows, x2 = log land, x3 = log feed, x4 = log labor

Regression Results

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=YIT     Mean                  =        11.57749
            Standard deviation    =          .64344
            No. of observations   =            1482  Degrees of freedom
Regression  Sum of Squares        =         584.056           4
Residual    Sum of Squares        =         29.0957        1477
Total       Sum of Squares        =         613.152        1481
            Standard error of e   =          .14035
Fit         R-squared             =          .95255  R-bar squared =    .95242
Model test  F[  4,  1477]         =      7412.18529  Prob F > F*   =    .00000
Diagnostic  Log likelihood        =       809.67609  Akaike I.C.   = -3.92381
            Restricted (b=0)      =     -1448.90834  Bayes  I.C.   = -3.90592
            Chi squared [   4]    =      4517.16885  Prob C2 > C2* =    .00000
Model was estimated on Aug 29, 2011 at 09:44:07 AM
--------+--------------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
     YIT|  Coefficient      Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
Constant|    11.5775***     .00365  3175.52  .0000     11.5703    11.5846
      X1|      .59518***    .01958    30.39  .0000      .55679     .63356
      X2|      .02305**     .01122     2.05  .0400      .00105     .04505
      X3|      .02319*      .01303     1.78  .0751     -.00235     .04873
      X4|      .45176***    .01078    41.89  .0000      .43062     .47290
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Distribution of OLS Residuals



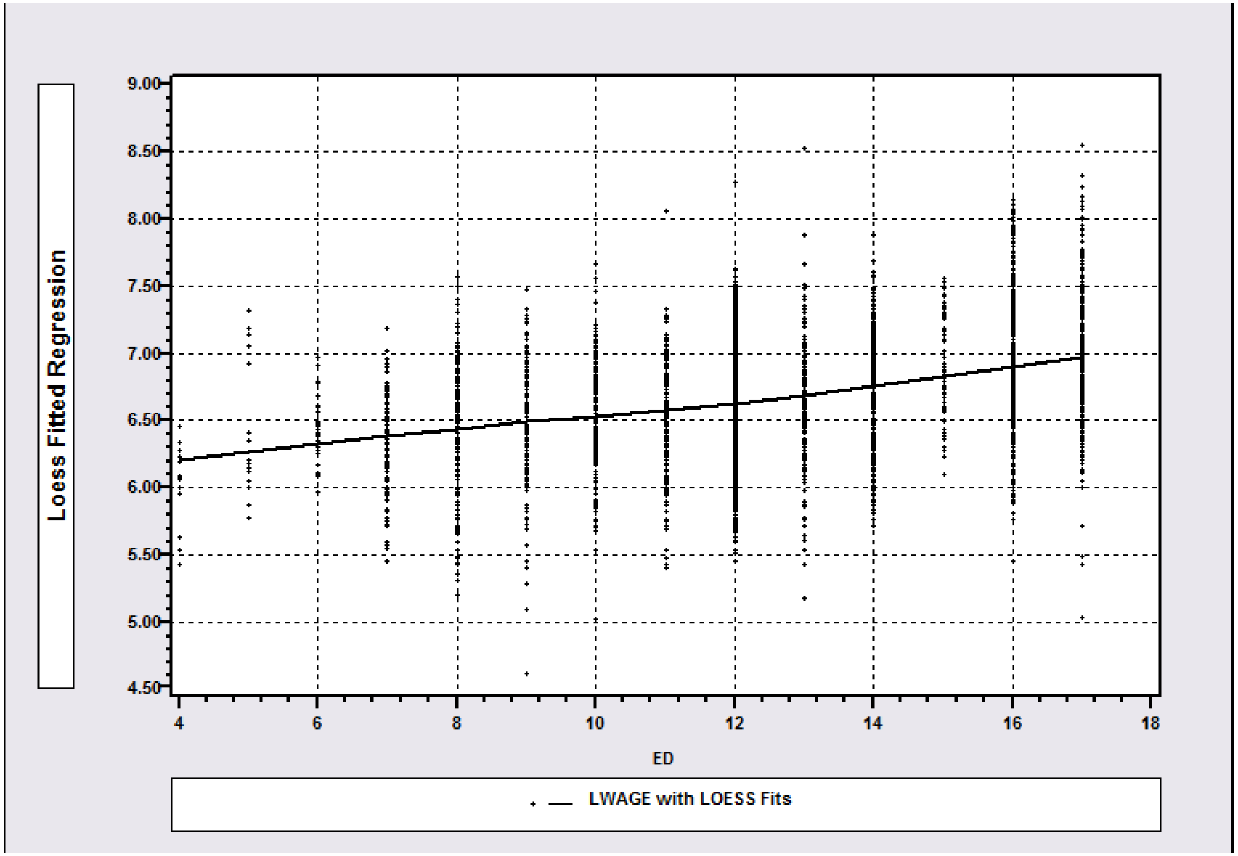
A Nonparametric Regression

- $y = \mu(x) + \varepsilon$
- Smoothing methods to approximate $\mu(x)$ at specific points, x^*
- For a particular x^*, $\mu(x^*) = \sum_i w_i(x^*|\mathbf{x})y_i$
  - E.g., for ols, $\mu(x^*) = a + bx^*$
  - $w_i = 1/n + (x_i - \bar{x}) / \Sigma_i (x_i - \bar{x})^2$
- We look for weighting scheme, local differences in relationship. OLS assumes a fixed slope, b.

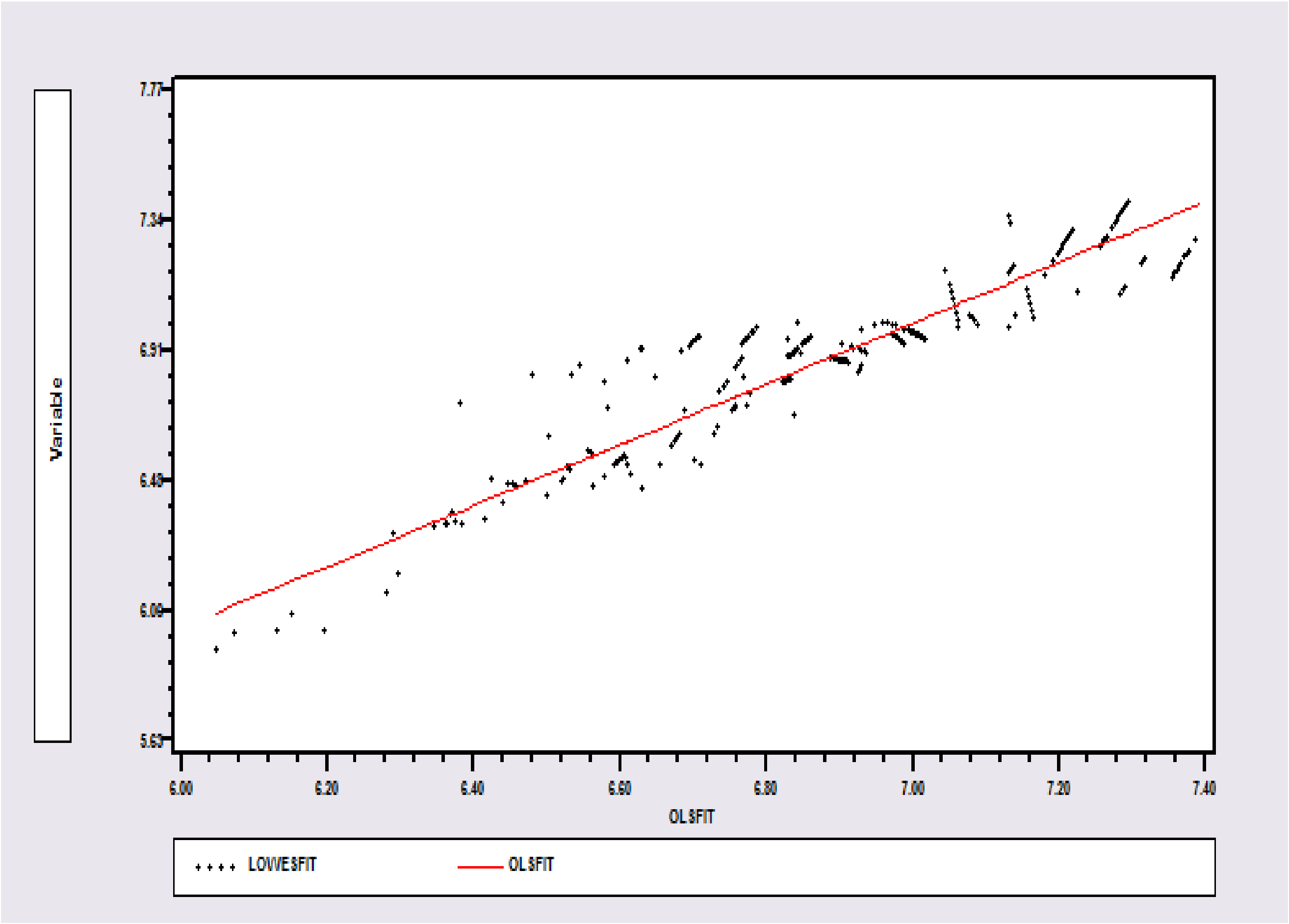
Nearest Neighbor Approach

- Define a neighborhood of x*. Points near get high weight, points far away get a small or zero weight
- Bandwidth, h defines the neighborhood:
  e.g., Silverman $h = .9 Min[s,(IQR/1.349)]/n^{.2}$
  Neighborhood is + or – h/2
- LOWESS weighting function: (tricube)
  $T_i = [1 - [Abs(x_i - x^*)/h]^3]^3$.
- Weight is $w_i = 1[Abs(x_i - x^*)/h < .5] * T_i$.

LOWESS Regression



OLS Vs. Lowess



Smooth Function: Kernel Regression

$$\hat{\mu}(x^* \mid \mathbf{x}, B) = \frac{\sum_{i=1}^{n} \frac{1}{B} K\left[\frac{x_i - x^*}{B}\right] y_i}{\sum_{i=1}^{n} \frac{1}{B} K\left[\frac{x_i - x^*}{B}\right]}$$

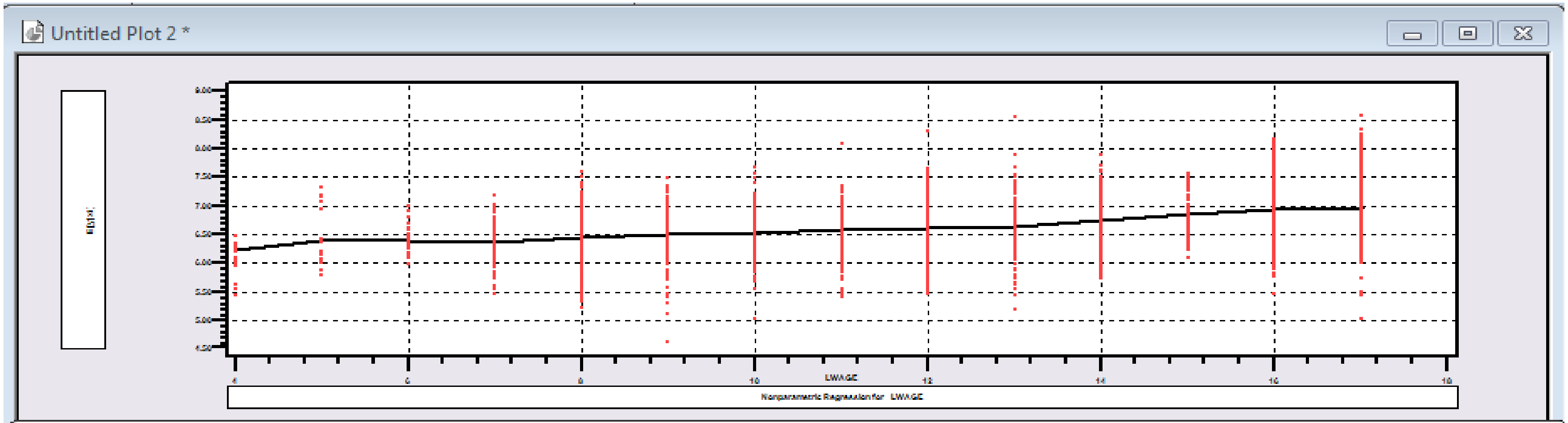
Kernel Functions:

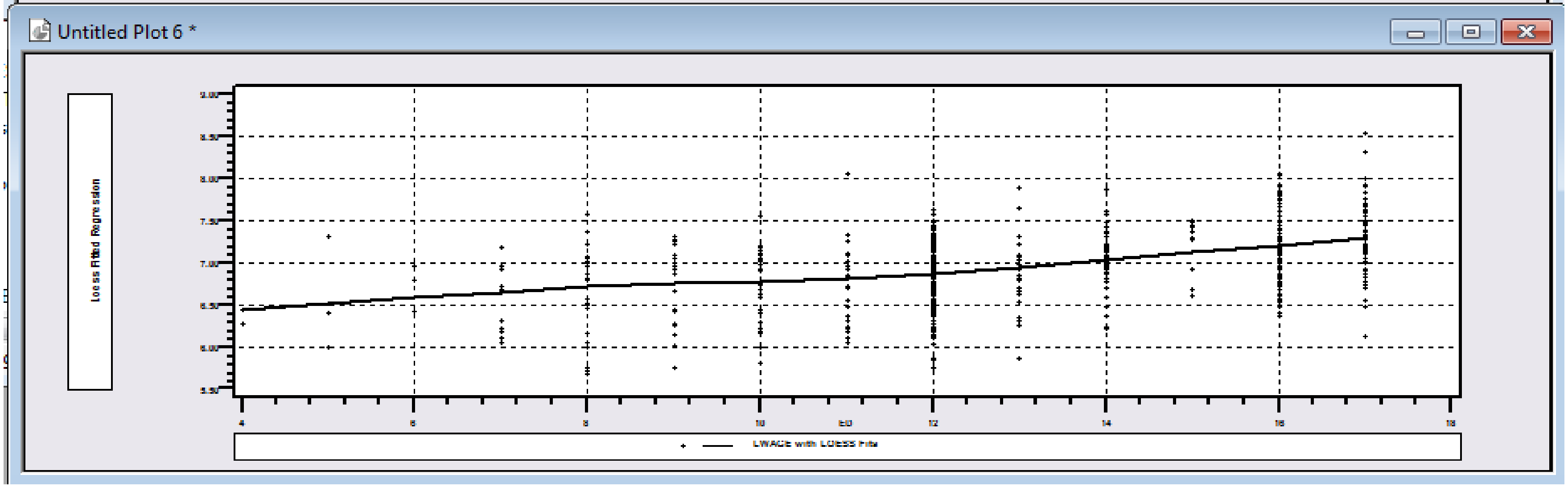
Normal: $K(t) = \phi(t)$

Logistic: $K(t) = \Lambda(t)[1-\Lambda(t)]$

Epanechnikov: $K(t) = .75(1 - .2t^2)/\sqrt{5}$, if $|t| \leq 5$ and 0 otherwise

Kernel Regression vs. Lowess (Lwage vs. Educ)





Locally Linear Regression

$$\mu(\mathbf{x}^*) = \beta(\mathbf{x}^*)'\mathbf{x}^*.$$

$$\beta(\mathbf{x}^*) = \left[\sum_{i=1}^{n} w_i(\mathbf{x}^*, \mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{n} w_i(\mathbf{x}^*, \mathbf{x}_i)\mathbf{x}_i y_i\right]$$

$$w_i(\mathbf{x}^*, \mathbf{x}_i) = K[(\mathbf{x}^* - \mathbf{x}_i)'(\mathbf{x}^* - \mathbf{x}_i), h]$$

OLS vs. LOWESS

```
-----------------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=LWAGE   Mean                 =        6.95074
            Standard deviation   =         .43840
            No. of observations  =            595  Degrees of freedom
Regression  Sum of Squares       =        38.4740            4
Residual    Sum of Squares       =        75.6913          590
Total       Sum of Squares       =        114.165          594
            Standard error of e  =         .35818
Fit         R-squared            =         .33700  R-bar squared =   .33251
Model test  F[  4,   590]        =       74.97447  Prob F > F*   =   .00000
Diagnostic  Log likelihood       =     -230.85370  Akaike I.C.   = -2.04509
            Restricted (b=0)     =     -353.12149  Bayes  I.C.   = -2.00821
            Chi squared [   4]   =      244.53559  Prob C2 > C2* =   .00000
Model was estimated on Aug 29, 2011 at 09:12:14 AM
--------+--------------------------------------------------------------------
        |                  Standard            Prob.      95% Confidence
   LWAGE|  Coefficient       Error        z    |z|>Z*        Interval
--------+--------------------------------------------------------------------
Constant|    5.57608***     .38522    14.48   .0000     4.82106   6.33109
      ED|     .07657***     .00548    13.97   .0000      .06583    .08732
     FEM|   -.44938***      .04704    -9.55   .0000    -.54157   -.35719
   UNION|     .10451***     .03239     3.23   .0013      .04103    .16798
  LOGWKS|     .10534        .09720     1.08   .2785    -.08517    .29585
--------+--------------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------------
```

Matrix - LOCLBETA

[595, 5] Cell: 0.305729

	1	2	3	4	5
1	4.78094	0.0473557	-0.541554	0.364504	0.360682
2	5.62988	0.0774372	-0.46378	0.102727	0.084603
3	6.97646	0.0351276	-0.365454	0.0817198	-0.134264
4	4.09146	0.0978699	-0.519258	0.212737	0.42492
5	4.24776	0.0951743	-0.408972	-0.0647883	0.387792
6	6.55862	0.048071	-0.414722	0.0847438	-0.066336
7	6.99921	0.0350738	-0.36576	0.0812928	-0.139938
8	5.80921	0.0136297	-0.583179	0.228573	0.198185
9	4.24905	0.0951543	-0.408957	-0.0648068	0.38754
10	4.25924	0.0949729	-0.408828	-0.0650166	0.385654