Do attractive people get paid more?

“Just remember that no matter how physically attractive or unattractive one may be, good grooming is still very important to one’s career.”

Felix Pretis (Oxford)
Econometrics
Oxford University, 2016
Econometrics: Computer Modelling

Felix Pretis

Programme for Economic Modelling
Oxford Martin School, University of Oxford

Lecture 1: Introduction to Econometric Software & Cross-Section Analysis
Aim of this Course

Aim:

- Introduce econometric modelling in practice
- Introduce OxMetrics/PcGive Software

By the end of the course:

- Able to build econometric models
- Evaluate output and test theories
- Use OxMetrics/PcGive to load, graph, model, data
Textbooks: no single text book. Useful:

  - Included in OxMetrics installation – “Help”
  - Freely available online: http://www.timberlake.co.uk/macroeconometrics.html

Lecture Notes & Lab Material online: http://www.felixpretis.org

Problem Set: to be covered in tutorial

Exam: Question possible (past paper from 2016)
1: Intro to Econometric Software & Cross-Section Regression
2: Micro-Econometrics: Limited Indep. Variable
3: Macro-Econometrics: Time Series
Motivation

- Economies high dimensional, interdependent, heterogeneous, and evolving: comprehensive specification of all events is impossible.

- Economic Theory
  - likely wrong and incomplete
  - **meaningless** without empirical support
  - Econometrics to discover new relationships from data
  - Econometrics can provide empirical support. . . or refutation.

- Require econometric software unless you really like doing matrix manipulation by hand.
### Structure of data

<table>
<thead>
<tr>
<th></th>
<th>wage</th>
<th>hours</th>
<th>IQ</th>
<th>kWh</th>
<th>educ</th>
<th>exper</th>
<th>tenure</th>
<th>age</th>
<th>married</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>769</td>
<td>40</td>
<td>93</td>
<td>35</td>
<td>12</td>
<td>11</td>
<td>2</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>808</td>
<td>50</td>
<td>119</td>
<td>41</td>
<td>18</td>
<td>11</td>
<td>16</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>825</td>
<td>40</td>
<td>108</td>
<td>46</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>650</td>
<td>40</td>
<td>96</td>
<td>32</td>
<td>12</td>
<td>13</td>
<td>7</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>562</td>
<td>40</td>
<td>74</td>
<td>27</td>
<td>11</td>
<td>14</td>
<td>5</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1400</td>
<td>40</td>
<td>116</td>
<td>43</td>
<td>16</td>
<td>14</td>
<td>2</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>600</td>
<td>40</td>
<td>91</td>
<td>24</td>
<td>10</td>
<td>13</td>
<td>0</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1081</td>
<td>40</td>
<td>114</td>
<td>50</td>
<td>18</td>
<td>8</td>
<td>14</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1154</td>
<td>45</td>
<td>111</td>
<td>37</td>
<td>15</td>
<td>13</td>
<td>1</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1000</td>
<td>40</td>
<td>95</td>
<td>44</td>
<td>12</td>
<td>16</td>
<td>16</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>930</td>
<td>43</td>
<td>132</td>
<td>44</td>
<td>18</td>
<td>8</td>
<td>13</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>921</td>
<td>38</td>
<td>102</td>
<td>45</td>
<td>14</td>
<td>9</td>
<td>11</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>900</td>
<td>45</td>
<td>125</td>
<td>40</td>
<td>15</td>
<td>4</td>
<td>3</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1318</td>
<td>38</td>
<td>119</td>
<td>24</td>
<td>16</td>
<td>7</td>
<td>2</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1792</td>
<td>40</td>
<td>118</td>
<td>47</td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>958</td>
<td>50</td>
<td>185</td>
<td>37</td>
<td>10</td>
<td>17</td>
<td>2</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>1360</td>
<td>45</td>
<td>189</td>
<td>39</td>
<td>15</td>
<td>6</td>
<td>9</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>850</td>
<td>40</td>
<td>72</td>
<td>36</td>
<td>11</td>
<td>19</td>
<td>10</td>
<td>38</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>830</td>
<td>44</td>
<td>105</td>
<td>29</td>
<td>14</td>
<td>4</td>
<td>7</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>471</td>
<td>44</td>
<td>181</td>
<td>34</td>
<td>12</td>
<td>13</td>
<td>7</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>1275</td>
<td>40</td>
<td>123</td>
<td>37</td>
<td>14</td>
<td>9</td>
<td>1</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>1615</td>
<td>50</td>
<td>113</td>
<td>49</td>
<td>16</td>
<td>10</td>
<td>4</td>
<td>36</td>
<td>1</td>
</tr>
</tbody>
</table>
Many options!

- **Menu-driven**
  - OxMetrics
  - STATA
  - EViews
  - ...

- **Simple batch programming language**
  - OxMetrics: batch-file
  - STATA: do-file
  - EViews: programme files
  - ...

- **Matrix-based programming languages**
  - R, Ox, Matlab, Python, C, Gauss,...
Putting it into practice: OxMetrics

- Computer program for working with economic data.
  - Front-end: Menu driven, easy to use.
  - But: Not restricted to menus: All underwritten by Ox Programming language.
- OxMetrics acts as an umbrella for many packages:
  - All have different names: PcGive, G@RCH, PcNaive, . . .
  - Different packages for different types of economic data.

OxMetrics Online Help

- Online OxMetrics help:
  http://www.pcgive.com/oxmetrics.html
- Online PcGive help:
  http://www.pcgive.com/pcgive/index.html
Download OxMetrics onto your own computers from:

http://www.doornik.com/download/Oxford/

- Free license for one year
- Download from within Oxford network
Open OxMetrics on your system.
→ **Layout of Window:** Structure, Results, Menus
What determines wages?

Education?
Ability
...Attractiveness?
Download the Data

- wages.in7

Load the Wages Data
- Open File
OxMetrics has great flexibility for plotting data.

- Huge range of possible types of plot. Easy to access.
- Multiple series on a set of axes, multiple sets of axes.
- Copy and paste works in wonderful ways...
- Can save in pdf format, ideal for including in \LaTeX{} or Word documents.
Always plot your Data

- as Scatter Plots, Histograms, etc.
- Modify the graphs!
Always plot your Data

- as *Scatter Plots*, *Histograms*, etc.
- Modify the graphs!
Economic theory often proposes objects of interest which are combinations:

- E.g. Average Propensity to Save: Savings/Income.
- E.g. Real exchange rate: \((\text{Spot rate} \times \text{Domestic prices})/\text{Foreign prices}\).
- Total Sales = Sales\(_1\) + Sales\(_2\)...
- \(\log(\text{Wages})\)

Hence we need to be able to create such variables in OxMetrics!

We use the Calculator tool:

- Model and Calculator.
- Alt+C.
Data transformations: Lags, logs, differences, percentage changes, ...

Other . . . : Extensive list of data transformations.

→ Do: Construct log(wages) variable
Understanding the Data & Building an Econometric Model:

1. Other: Descriptive Statistics (Summary Statistics etc.)
   - Task: Create summary statistics for "Wages" dataset

2. Cross Section Regression (Regression Model)
   - Build simple model of log(wages): regress log(wages) on education and an intercept
Create summary statistics:

Means, standard deviations and correlations
The sample is: 1 – 935 (935 observations and 6 variables)

Means

<table>
<thead>
<tr>
<th></th>
<th>wage</th>
<th>hours</th>
<th>IQ</th>
<th>KWW</th>
<th>educ</th>
<th>exper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>957.95</td>
<td>43.929</td>
<td>101.28</td>
<td>35.744</td>
<td>13.468</td>
<td>11.564</td>
</tr>
</tbody>
</table>

Standard deviations (using T-1)

<table>
<thead>
<tr>
<th></th>
<th>wage</th>
<th>hours</th>
<th>IQ</th>
<th>KWW</th>
<th>educ</th>
<th>exper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std Dev</td>
<td>404.36</td>
<td>7.2243</td>
<td>15.053</td>
<td>7.6388</td>
<td>2.1967</td>
<td>4.3746</td>
</tr>
</tbody>
</table>

Correlation matrix:

<table>
<thead>
<tr>
<th></th>
<th>wage</th>
<th>hours</th>
<th>IQ</th>
<th>KWW</th>
<th>educ</th>
<th>exper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage</td>
<td>1.0000</td>
<td>-0.0095043</td>
<td>0.30909</td>
<td>0.32613</td>
<td>0.32711</td>
<td>0.0021897</td>
</tr>
<tr>
<td>Hours</td>
<td>-0.0095043</td>
<td>1.0000</td>
<td>0.073839</td>
<td>0.11389</td>
<td>0.091009</td>
<td>-0.062126</td>
</tr>
<tr>
<td>IQ</td>
<td>0.30909</td>
<td>0.073839</td>
<td>1.0000</td>
<td>0.41352</td>
<td>0.51570</td>
<td>-0.22491</td>
</tr>
<tr>
<td>KWW</td>
<td>0.32613</td>
<td>0.11389</td>
<td>0.41352</td>
<td>1.0000</td>
<td>0.38813</td>
<td>0.017452</td>
</tr>
<tr>
<td>Educ</td>
<td>0.32711</td>
<td>0.091009</td>
<td>0.51570</td>
<td>0.38813</td>
<td>1.0000</td>
<td>-0.45557</td>
</tr>
<tr>
<td>Exper</td>
<td>0.0021897</td>
<td>-0.062126</td>
<td>-0.22491</td>
<td>0.017452</td>
<td>-0.45557</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Interpreting Regression Output

EQ( 1) Modelling Lwage by OLS-CS
The dataset is: wages.in7
The estimation sample is: 1 - 935

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.97306</td>
<td>0.08137</td>
<td>73.4</td>
<td>0.0000</td>
</tr>
<tr>
<td>educ</td>
<td>0.0598392</td>
<td>0.005963</td>
<td>10.0</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

| sigma       | 0.40032   |         |        |          |
| RSS         | 149.518582|         |        |          |
| F(1,933) = 100.7 [0.000]** |         |        |        |          |
| Adj.R^2     | 0.0964494 |         |        |          |
| log-likelihood | -469.721 |        |        |          |
| no. of observations | 935 | | | |
| mean(Lwage) | 6.779    | se(Lwage) | 0.421144 |
Omitted variable-bias:

- **Our model**: \( L(wage)_i = \beta_0 + \beta_1 \text{Educ}_i + v_i \)
- **What if**: Data Generating Process “Truth”=
  \( L(wage)_i = \beta_0 + \beta_1 \text{Educ}_i + \beta_2 \text{Ability}_i + \epsilon_i \)

Then \( \hat{\beta}_1 \) biased unless \( \beta_1 = 0 \) or Educ. & Ability are uncorrelated.

Demonstration: **Monte Carlo Simulation**. Suppose

- \( \beta_1 = 1 \) (Education)
- \( \beta_2 = 4 \) (Ability, IQ)

With:

- A: Corr(Education, Ability)=0.5
- B: Corr(Education, Ability)=0
Concerns about omitted variables

Start **general**, then reduce to **specific**.

**Determinants of Wages** – build a more general regression model, including:

- IQ (proxy for ability)
- Experience
- Age
- Race (black)
### More general model

The estimation sample is: 1 - 935

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.04194</td>
<td>0.1719</td>
<td>29.3</td>
<td>0.0000</td>
</tr>
<tr>
<td>educ</td>
<td>0.0517412</td>
<td>0.007549</td>
<td>6.85</td>
<td>0.0000</td>
</tr>
<tr>
<td>IQ</td>
<td>0.00454855</td>
<td>0.001037</td>
<td>4.39</td>
<td>0.0000</td>
</tr>
<tr>
<td>exper</td>
<td>0.0132632</td>
<td>0.003844</td>
<td>3.45</td>
<td>0.0006</td>
</tr>
<tr>
<td>age</td>
<td>0.0134906</td>
<td>0.004827</td>
<td>2.79</td>
<td>0.0053</td>
</tr>
<tr>
<td>black</td>
<td>-0.156964</td>
<td>0.04056</td>
<td>-3.87</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

sigma | 0.381616 | RSS 135.290875 |
R\(^2\) | 0.183304 | F(5, 929) = 41.7 [0.000]** |
Adj.R\(^2\) | 0.178908 | log-likelihood -422.974 |
no. of observations | 935 | no. of parameters 6 |
mean(Lwage) | 6.779 | se(Lwage) 0.421144 |
Testing Hypotheses on Model Parameters:

- Exclusion restrictions: Test - Exclusion Restrictions
- General restrictions: Test - General Restrictions

Do:

- Excluding Age & Black?
- Coefficient on Exper = Coefficient on Age?
- Coefficient on Educ = 0.05?
Testing for Heteroskedasticity:

- White Test for Heteroskedasticity (without cross-products)

\[ \hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_1^2 + \delta_4 x_2^2 \cdots + \text{error} \quad (1) \]

Test: \( \delta_1 = 0, \delta_2 = 0, \ldots \)

Need to:

- Store residuals from regression \( \hat{u} \)
- Generate new variables: \( \hat{u}^2, x_1^2, \ldots \)
- Estimate White model & test exclusion
Testing for Heteroskedasticity

Need to:

- Store residuals from regression \( \hat{u} \) – *Test - store Residuals*...
- Generate new variables: \( \hat{u}^2, x_1^2, \ldots \) – *Calculator Function*
- Estimate White test model & test exclusion – as before

Test for excluding:

[0] = educ  
[1] = IQ  
[2] = exper  
[3] = age  
[4] = black  
[5] = educ_sq  
[6] = IQ_sq  
[7] = exper_sq  
[8] = age_sq

Subset \( F(9, 925) = 0.55046 \) [0.8378]
Testing for Heteroskedasticity

Luckily OxMetrics reports many diagnostic tests automatically:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.04194</td>
<td>0.1719</td>
<td>29.3</td>
<td>0.0000</td>
</tr>
<tr>
<td>educ</td>
<td>0.0517412</td>
<td>0.007549</td>
<td>6.85</td>
<td>0.0000</td>
</tr>
<tr>
<td>IQ</td>
<td>0.00454855</td>
<td>0.001037</td>
<td>4.39</td>
<td>0.0000</td>
</tr>
<tr>
<td>exper</td>
<td>0.0132632</td>
<td>0.003844</td>
<td>3.45</td>
<td>0.0006</td>
</tr>
<tr>
<td>age</td>
<td>0.0134906</td>
<td>0.004827</td>
<td>2.79</td>
<td>0.0053</td>
</tr>
<tr>
<td>black</td>
<td>-0.156964</td>
<td>0.04056</td>
<td>-3.87</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

| sigma       | 0.381616   | RSS     | 135.290875 |
| R^2         | 0.183304   | F(5,929) = 41.7 [0.000]** |
| Adj.R^2     | 0.178908   | log-likelihood -422.974 |
| no. of observations | 935 |
| mean(Lwage) | 6.779      | se(Lwage) | 0.421144 |

Normality test: Chi^2(2) = 28.678 [0.0000]**
Hetero test: F(9,925) = 0.55046 [0.8378]
Hetero-X test: F(15,919) = 0.59513 [0.8803]
RESET23 test: F(2,927) = 0.58041 [0.5599]

**Compare:** Subset F(9,925) = 0.55046 [0.8378]
Luckily OxMetrics reports many diagnostic tests automatically:

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Std.Error</th>
<th>t-value</th>
<th>t-prob</th>
<th>Part.R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.04194</td>
<td>0.1719</td>
<td>29.3</td>
<td>0.0000</td>
</tr>
<tr>
<td>educ</td>
<td>0.0517412</td>
<td>0.007549</td>
<td>6.85</td>
<td>0.0000</td>
</tr>
<tr>
<td>IQ</td>
<td>0.00454855</td>
<td>0.001037</td>
<td>4.39</td>
<td>0.0000</td>
</tr>
<tr>
<td>exper</td>
<td>0.0132632</td>
<td>0.003844</td>
<td>3.45</td>
<td>0.0006</td>
</tr>
<tr>
<td>age</td>
<td>0.0134906</td>
<td>0.004827</td>
<td>2.79</td>
<td>0.0053</td>
</tr>
<tr>
<td>black</td>
<td>-0.156964</td>
<td>0.04056</td>
<td>-3.87</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

\[
\text{sigma} \quad 0.381616 \quad \text{RSS} \quad 135.290875
\]
\[
\text{R}^2 \quad 0.183304 \quad F(5,929) = 41.7 \quad [0.000]\quad **
\]
\[
\text{Adj.R}^2 \quad 0.178908 \quad \log\text{-likelihood} \quad -422.974
\]

<table>
<thead>
<tr>
<th>no. of observations</th>
<th>935</th>
<th>no. of parameters</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(Lwage)</td>
<td>6.779</td>
<td>se(Lwage)</td>
<td>0.421144</td>
</tr>
</tbody>
</table>

Normality test: \( \text{Chi}^2(2) = 28.678 \quad [0.0000]\quad ** \)
Hetero test: \( F(9,925) = 0.55046 \quad [0.8378] \)
Hetero-X test: \( F(15,919) = 0.59513 \quad [0.8803] \)
RESET23 test: \( F(2,927) = 0.58041 \quad [0.5599] \)

What about other diagnostic tests?
Education and experience matters, but what about attractiveness?
OxMetrics & PcGive Exercise: Attractiveness and Wages

- Load Data: "attractiveness.in7"
  - "looks": variable from 1-5 scored looks
  - "attr_belavg": dummy variable below average attractiveness
  - "attr_abvavg": dummy variable above average attractiveness

- Graph Data

- Build model of log(Wages)

*American Economic Review.* Vol. 84 (5) (goo.gl/3bP3IA)
Answering the following questions:

1. What measure of attractiveness is most sensible to include?
2. What proportion of the population are rated as above/below average?
3. Is there a significant difference in above/below average attractiveness between men and women? (Hint: think about a regression model that would allow you to test this!)
4. Does attractiveness have an effect on wages? Quantify any effect you find.
5. Is the effect different for men/women?
6. Does the effect of attractiveness change when controlling for other variables? What about other interaction effects?
7. Is your model well-specified?
8. Is there evidence for Heteroskedasticity in your models?
9. What about other diagnostic criteria (Normality etc.)?