

BUEC 333:  
An Introduction to  
Introduction to Econometrics

*Krishna Pendakur, Economics, SFU*

# What is Econometrics?

- *Econometrics* is the application of economic theory and statistical methods to analyze economic data.
- It's both a science and an art.
  - There's always a disconnect between economic theory and the real world. Creatively bridging this gap to apply economic theory to real-world data is an art.
  - Economic data are rarely "perfect" (experimental) -- sometimes requires some finesse to get believable results out of them.
  - Economic theory and statistics are grounded in mathematics. Direct application of either is scientific.

# Econometric Questions

- *Magnitudes and Tests*
- Does reducing class size improve the quality of education?
  - Common sense tells us that students get more attention in smaller classes, and therefore probably learn more. **But how much more?** Reducing class size costs money (need more teachers). Does the improved quality of education justify the expense? Here, the *magnitude* of improvement matters.
- What will be the value of the S&P 500 one year from today?
  - Knowing (with certainty) the answer to this question would be worth a lot.
  - A good (accurate) estimate is also valuable.
- Do Immigrants Face Discrimination in the Labour Market?
  - This is a yes/no question. We can use statistics to *test* whether or not they face disparity.

# So What?

- Knowing some basic econometrics can be very lucrative.
  - Lots of jobs, some pay very well.
- Knowing your way around some econometric software can too.
- Useful for future study (4th year courses -- you'll see/read quite a bit of empirical work, and maybe do some).
- Even if you end up doing something completely different, it's useful for understanding the world around you.
  - Managers need to understand where estimates come from and what they tell you (and what they don't!)
  - Helpful for understanding real-world statistics (polls, press about empirical studies, "facts" and arguments, etc.)
  - Helpful for understanding finance (alphas, betas,  $R^2$ , etc.)
- Good Policy Uses Econometric Analysis
  - Magnitudes Matter
  - Causation Matters

# BUEC 333 Deliverables...

- In this class, you will learn the statistical methods you would use to answer these questions (and many more ...)
- You will learn how to assess whether a particular estimate/analysis is any good.
- You will learn to use a specific statistical software package (EViews)
- You will get some practice working with real data.

# Today

- Correlation
- Magnitudes Matter
- Statistical Precision Matters
- One Dimension Isn't Enough
- Regression
- Causation
- Instruments, Natural Experiments and Field Experiments

# Some Data

- Long Example: Do immigrants earn less?
  - How? Why?
- 2006 Census public-use *microdata*, sample of from Census long forms.
- 3% sample of population of BC, about 100,000 observations of individuals
- Earnings from Wages and Salaries (ie., working for other people) in 2005

# What is Microdata? (Excel)

| ABOID | AGEGRP | AGEIMM | ATTSCH | BFNMEMB | CFINC | CFINC_AT | CFINEF | CFSIZE | CFSTAT |
|-------|--------|--------|--------|---------|-------|----------|--------|--------|--------|
| 6     | 1      | 99     | 9      | 1       | 13    | 12       | 1      | 3      | 8      |
| 6     | 5      | 2      | 9      | 1       | 12    | 12       | 1      | 5      | 7      |
| 6     | 11     | 8      | 1      | 1       | 10    | 10       | 2      | 2      | 1      |
| 6     | 18     | 10     | 1      | 1       | 99    | 99       | 9      | 1      | 12     |
| 6     | 8      | 99     | 1      | 1       | 27    | 25       | 1      | 4      | 8      |
| 6     | 6      | 99     | 2      | 1       | 28    | 26       | 1      | 5      | 7      |
| 6     | 11     | 99     | 1      | 1       | 99    | 99       | 9      | 1      | 12     |
| 1     | 2      | 99     | 9      | 2       | 11    | 10       | 1      | 3      | 10     |
| 6     | 18     | 13     | 1      | 1       | 23    | 23       | 1      | 4      | 1      |
| 6     | 1      | 99     | 9      | 1       | 4     | 4        | 1      | 5      | 7      |
| 6     | 11     | 7      | 1      | 1       | 8     | 8        | 1      | 3      | 6      |
| 6     | 12     | 8      | 1      | 1       | 15    | 14       | 1      | 3      | 1      |
| 6     | 14     | 8      | 1      | 1       | 14    | 13       | 1      | 2      | 2      |
| 6     | 12     | 99     | 1      | 1       | 22    | 21       | 1      | 4      | 2      |
| 6     | 16     | 3      | 1      | 1       | 99    | 99       | 9      | 1      | 12     |
| 6     | 6      | 99     | 2      | 1       | 28    | 26       | 1      | 4      | 7      |
| 6     | 14     | 9      | 1      | 1       | 88    | 88       | 1      | 2      | 1      |
| 6     | 5      | 99     | 9      | 1       | 9     | 9        | 1      | 3      | 10     |
| 6     | 14     | 8      | 1      | 1       | 10    | 10       | 2      | 3      | 1      |
| 6     | 11     | 5      | 1      | 1       | 19    | 18       | 2      | 5      | 1      |

# What is Microdata?

- But, there are 56,530 rows and 124 columns
- A bit cumbersome
- Need to summarize this information
  - Cross-Tabs
  - Tables of Mean
  - Regressions

# Correlation

- When one number is high, do we expect another number to be high (or low)?
- When a person is an immigrant, are their earnings lower?
- Need to map out (at least) immigration status and earnings

# The *Cross-Tab*

*Cross-Tabs* show the numbers of observations (rows of the data) that have values in different ranges for different variables. They show how the data are clumped, and how the data vary with each other.

| Immigrant status      | groupwAGES |         |             |        | Total  |
|-----------------------|------------|---------|-------------|--------|--------|
|                       | zero       | 1-\$20k | \$20K-\$50k | \$50k+ |        |
| Temporary Residents   | 527        | 239     | 115         | 72     | 953    |
| Canadian-born         | 7,289      | 6,212   | 6,520       | 4,881  | 24,902 |
| Foreign-born Perm Res | 8,814      | 5,132   | 4,878       | 2,355  | 21,179 |
| Total                 | 16,630     | 11,583  | 11,513      | 7,308  | 47,034 |

- Seems crazy to talk about the earnings of children, and the real elderly, and people who arrived after the 2005 income year
- So, drop
  - persons under 25
  - persons over 69
  - immigrants who arrived in 2005 or 2006

| Immigrant status      | groupWAGES |         |             |        | Total  |
|-----------------------|------------|---------|-------------|--------|--------|
|                       | Zero       | 1-\$20k | \$20K-\$50k | \$50k+ |        |
| Temporary Residents   | 274        | 162     | 103         | 69     | 608    |
| Canadian-born         | 3,962      | 3,125   | 5,823       | 4,820  | 17,730 |
| Foreign-born Perm Res | 5,206      | 3,608   | 4,573       | 2,304  | 15,691 |
| Total                 | 9,442      | 6,895   | 10,499      | 7,193  | 34,029 |

# Cross-Tabs with Column %s

| Immigrant status      | groupWAGES |         |             |        | Total  |
|-----------------------|------------|---------|-------------|--------|--------|
|                       | zero       | 1-\$20k | \$20K-\$50k | \$50k+ |        |
| Temporary Residents   | 2.90       | 2.35    | 0.98        | 0.96   | 1.79   |
| Canadian-born         | 41.96      | 45.32   | 55.46       | 67.01  | 52.10  |
| Foreign-born Perm Res | 55.14      | 52.33   | 43.56       | 32.03  | 46.11  |
| Total                 | 100.00     | 100.00  | 100.00      | 100.00 | 100.00 |

# Cross-Tabs with Row %s

| Immigrant status      | groupwAGES |         |             |        | Total  |
|-----------------------|------------|---------|-------------|--------|--------|
|                       | zero       | 1-\$20k | \$20K-\$50k | \$50k+ |        |
| Temporary Residents   | 45.07      | 26.64   | 16.94       | 11.35  | 100.00 |
| Canadian-born         | 22.35      | 17.63   | 32.84       | 27.19  | 100.00 |
| Foreign-born Perm Res | 33.18      | 22.99   | 29.14       | 14.68  | 100.00 |
| Total                 | 27.75      | 20.26   | 30.85       | 21.14  | 100.00 |

# Do Immigrants Earn Less?

- Temporary Residents have **way** more zero-earners, and foreign-born have somewhat more zero-earners
- Foreign-born permanent residents more low-earners and less high-earners than Canadian-borns. The pattern is even stronger for temporary residents.

# Integration Takes Time

- Immigrants are heterogeneous
- Some arrived long ago, some very recently
- Can we account for this in a Cross-Tab?

# Do Immigrants Earn Less?

| Immigrant Cohort | groupwAGES |         |             |        | Total  |
|------------------|------------|---------|-------------|--------|--------|
|                  | zero       | 1-\$20k | \$20K-\$50k | \$50k+ |        |
| Before 1950      | 37.88      | 19.70   | 19.70       | 22.73  | 100.00 |
| 1950s            | 40.84      | 16.40   | 20.77       | 21.99  | 100.00 |
| 1960s            | 40.87      | 14.92   | 24.44       | 19.76  | 100.00 |
| 1970s            | 29.79      | 16.16   | 29.94       | 24.11  | 100.00 |
| 1980s            | 29.51      | 19.50   | 33.49       | 17.51  | 100.00 |
| 1990s            | 34.77      | 25.92   | 29.23       | 10.07  | 100.00 |
| 2000s            | 33.66      | 32.07   | 27.28       | 6.98   | 100.00 |
| Cdn-Born         | 22.35      | 17.63   | 32.84       | 27.19  | 100.00 |
| Temporary        | 45.07      | 26.64   | 16.94       | 11.35  | 100.00 |
| Total            | 27.80      | 20.11   | 30.81       | 21.28  | 100.00 |

# Table of Means

- Only 7% of recent immigrants have high earnings, compared with 27% of Cdn-born.
- But, 24% of immigrants from the 1970s have high earnings.
- This is getting clunky: how else can we summarize these data?
- *A Table of Means* gives the average value of some variable for values or ranges of another variable

# Table of Means

| Immigrant Cohort | N(WAGES) | mean(WAGES) |
|------------------|----------|-------------|
| Before 1950      | 66       | 27561       |
| 1950s            | 573      | 30682       |
| 1960s            | 1,260    | 31760       |
| 1970s            | 2,679    | 33365       |
| 1980s            | 2,416    | 27478       |
| 1990s            | 5,412    | 20283       |
| 2000s            | 2,463    | 17103       |
| Cdn-Born         | 17,730   | 38317       |
| Temporary        | 608      | 21737       |

# Table of Means

- Those 2000s arrivals have it bad: \$20,000 lower earnings than Canadian-borns.
- 1990s arrivals have similarly low earnings.
- But, 1970s arrivals have earnings about \$5000 lower than Canadian-born.

# Variation

- Used 3% of Greater Vancouver households.
- The mean is not the mean---it is an *estimate* of the mean.
- There's dispersion around the mean. Lots of people have income far from the mean.
- *Standard Deviation (SD)* characterises how far people are from the mean
  - *SD=the average squared distance from the mean*

# Table of Means and SDs

| Immigrant Cohort | N(WAGES) | mean(WAGES) | sd(WAGES) |
|------------------|----------|-------------|-----------|
| Before 1950      | 66       | 27561       | 38380     |
| 1950s            | 573      | 30682       | 58677     |
| 1960s            | 1,260    | 31760       | 65581     |
| 1970s            | 2,679    | 33365       | 49297     |
| 1980s            | 2,416    | 27478       | 34858     |
| 1990s            | 5,412    | 20283       | 31889     |
| 2000s            | 2,463    | 17103       | 24848     |
| Cdn-Born         | 17,730   | 38317       | 56307     |
| Temporary        | 608      | 21737       | 57811     |

- Immigrants arriving 1961-1970 and 1971-1980 earn almost as much as Canadian-born people
- Those arriving after seem to earn less and less
- But, are these differences statistically significant---that is, are we sure that they are there?

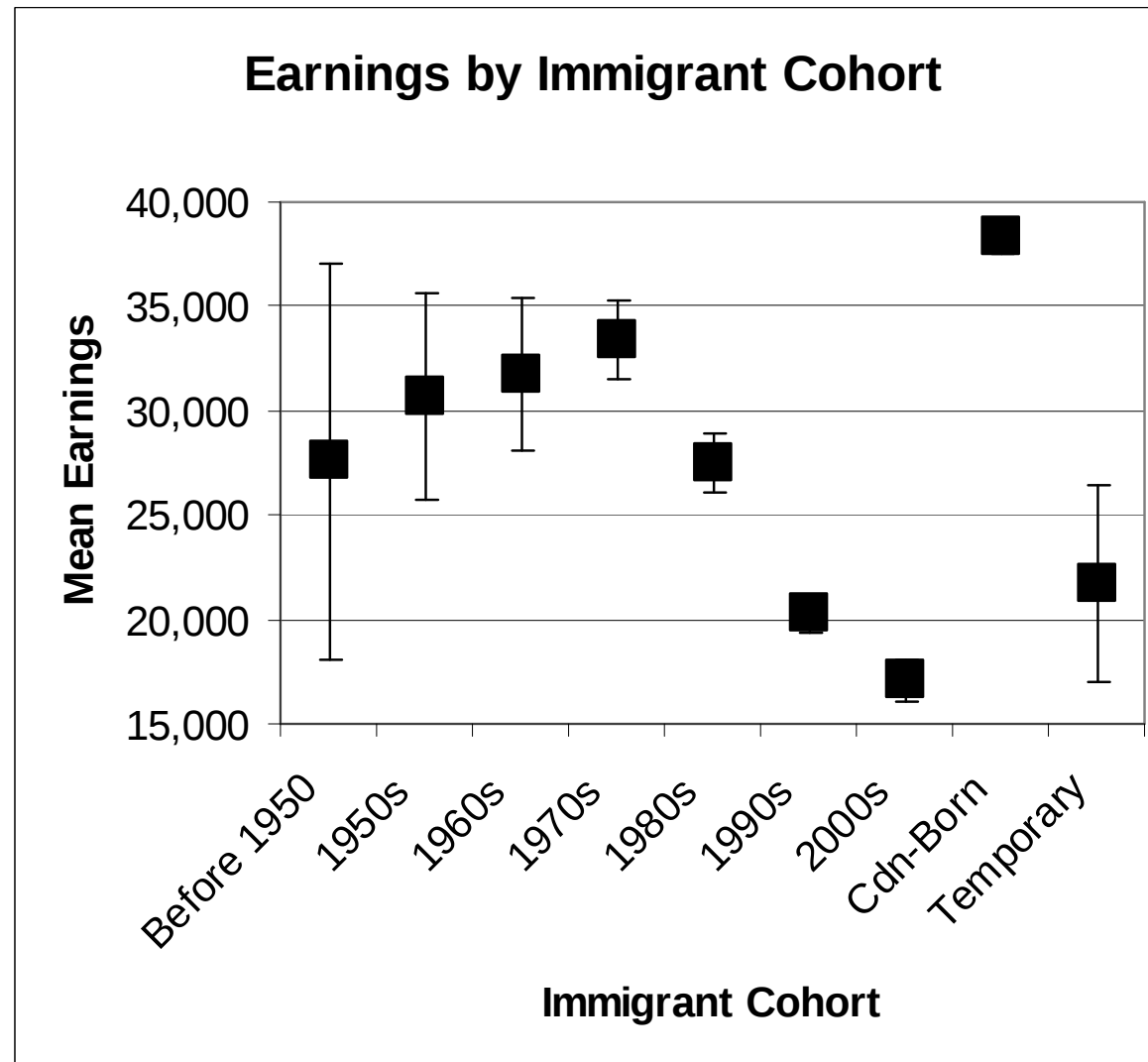
# Precision

- The *precision* of an estimated mean depends on
  - How much data you have ( $N$ )
  - How much dispersion there is ( $sd$ )
- *Standard Error (SE)*
  - equals the standard deviation divided by the square root of  $N$ :  $SE = SD/\sqrt{N}$
- *95% Confidence Interval (CI)*
  - gives a range such that we're 95% sure that the true mean lies inside. These are roughly 4 SEs wide (2 above, and 2 below, the estimated value)

# Table of Means, SDs and SEs

| Immigrant Cohort | N(WAGES) | mean(WAGES) | sd(WAGES) | sem(WAGES) |
|------------------|----------|-------------|-----------|------------|
| Before 1950      | 66       | 27561       | 38380     | 4724       |
| 1950s            | 573      | 30682       | 58677     | 2451       |
| 1960s            | 1,260    | 31760       | 65581     | 1848       |
| 1970s            | 2,679    | 33365       | 49297     | 952        |
| 1980s            | 2,416    | 27478       | 34858     | 709        |
| 1990s            | 5,412    | 20283       | 31889     | 433        |
| 2000s            | 2,463    | 17103       | 24848     | 501        |
| Cdn-Born         | 17,730   | 38317       | 56307     | 423        |
| Temporary        | 608      | 21737       | 57811     | 2345       |

## Another Way To Present the Means and Standard Error



# So, Do Immigrants Earn Less?

- The 95% confidence interval for the mean earnings of a Canadian-born person aged 25-69 is entirely above that of the mean earnings of every immigrant cohort.
- A *test* considers the probability that some relationship is true.
- Can we test whether or not immigrants earn less? Given the probabilities, it is highly improbable that they earn the same as the Canadian-born. So, we reject the hypothesis that they earn the same.

# Age and Arrival Cohort

- Immigrants arriving before 1961 and after 1980 earn less than Canadian-born workers.
- But, those arriving before 1961 are old.
- We know old people have low earnings, because they quit working.
- Those arriving between 1961 and 1980 are in their prime earning years.
- Is this all about age?

# Immigrants Age and Earnings

| Immigrant Cohort | Age Group |       |       |       |       |
|------------------|-----------|-------|-------|-------|-------|
|                  | 25-34     | 35-44 | 45-54 | 55-64 | 65-69 |
| Before 1950      |           |       |       | 31765 | 13267 |
| 1950s            |           |       | 49199 | 36359 | 7537  |
| 1960s            |           | 53708 | 42967 | 28865 | 10187 |
| 1970s            | 37704     | 41336 | 33789 | 33291 | 6876  |
| 1980s            | 31648     | 31057 | 30124 | 19169 | 4008  |
| 1990s            | 20317     | 25139 | 22777 | 10365 | 2117  |
| 2000s            | 18793     | 19564 | 15538 | 8810  | 2896  |
| Cdn-Born         | 31552     | 43511 | 47176 | 35218 | 10337 |
| Temporary        | 16820     | 26950 | 38551 | 13000 | 1333  |

# Do Immigrants Earn Less?

- Now that we consider people who are the same age, the disparities look smaller.
- Immigrants aged 25-34 who arrived in the 1970s and 1980 have average earnings **higher** than their Canadian-born counterparts.
- But, a lot of age-arrival year groups earn less.

# Controlling for Age

- When we *control* for a variable  $x$ , we ask “what is the difference in  $y$  for a given value of  $x$ ?”
- When you look **within** age categories, there’s less disparity.
- Thus, *controlling* for age reduces the amount of disparity we see between immigrants and Canadian-born people

# Controlling for Education

- Immigrants are more educated than Canadian-born people. Shouldn't they earn more?

| Immigrant Cohort | Highest Level of Schooling |       |         |        |         | Total  |
|------------------|----------------------------|-------|---------|--------|---------|--------|
|                  | LT HS                      | HS    | Some PS | BA/BSc | Post BA |        |
| Before 1950      | 13.64                      | 16.67 | 36.36   | 1.52   | 31.82   | 100.00 |
| 1950s            | 16.75                      | 27.92 | 30.89   | 4.89   | 19.55   | 100.00 |
| 1960s            | 13.67                      | 24.09 | 33.07   | 6.36   | 22.81   | 100.00 |
| 1970s            | 13.14                      | 22.10 | 30.20   | 7.61   | 26.95   | 100.00 |
| 1980s            | 17.00                      | 23.91 | 24.70   | 8.81   | 25.57   | 100.00 |
| 1990s            | 15.22                      | 23.35 | 19.93   | 9.29   | 32.20   | 100.00 |
| 2000s            | 10.68                      | 17.63 | 13.89   | 9.42   | 48.38   | 100.00 |
| Cdn-Born         | 9.21                       | 26.27 | 32.47   | 5.36   | 26.69   | 100.00 |
| Temporary        | 4.94                       | 14.99 | 16.97   | 8.90   | 54.20   | 100.00 |
| Total            | 11.41                      | 24.37 | 28.02   | 6.82   | 29.38   | 100.00 |

# Mean Earnings by Education

| Immigrant Cohort | Highest Level of Schooling |       |         |        |         |
|------------------|----------------------------|-------|---------|--------|---------|
|                  | LT HS                      | HS    | Some PS | BA/BSC | Post BA |
| Before 1950      | 10222                      | 11000 | 43167   | 12000  | 26571   |
| 1950s            | 17343                      | 26142 | 29176   | 34143  | 49405   |
| 1960s            | 13041                      | 26711 | 34918   | 29211  | 44817   |
| 1970s            | 15679                      | 24290 | 34611   | 32363  | 48300   |
| 1980s            | 11915                      | 20463 | 29960   | 29356  | 41335   |
| 1990s            | 9438                       | 14238 | 21435   | 20271  | 29106   |
| 2000s            | 8502                       | 12737 | 15944   | 13242  | 21692   |
| Cdn-Born         | 20218                      | 33702 | 35447   | 40512  | 52129   |
| Temporary        | 11533                      | 13571 | 18226   | 12037  | 27577   |

# Controlling for Education

- Unlike when we controlled for age, when we control for education, the disparity gets **bigger**.
- So, is the disparity big or small? Should we control for age or education?
- You could make an even bigger table of means, or...we could run a regression.

# Regression Analysis

- Tool to look at differences along many dimensions (ie., along age, education and immigration cohort all at the same time)
- Modeling Assumption: the effects of all these things can be summarised by **adding up** a bunch of individual effects.
- Then, assessing any one of them is easy: its effect on the total is isolated.

# Analysing Regressions

- Cdn-Born, HS, aged 35-44 has estimated earnings of \$29,168 (top number)
- With a PhD (but still Cdn-Born aged 35-44), it would be \$23,459 higher, equalling \$52,627
- So, you can isolate the effect of having a PhD
- You can also isolate the effect of immigration arrival year

# Regression Estimates

| Variable                 | Estimate | <i>Std Err</i> |
|--------------------------|----------|----------------|
| <b>Cdn-born 35-44 HS</b> | 37629    | 742            |
| Before 1950              | -3662    | 5999           |
| 1950s                    | 952      | 2101           |
| 1960s                    | -1538    | 1446           |
| 1970s                    | -4382    | 1020           |
| 1980s                    | -10952   | 1057           |
| 1990s                    | -18928   | 760            |
| 2000s                    | -25217   | 1059           |
| Temporary                | -20328   | 2017           |

| Variable | Estimate | <i>Std Err</i> |
|----------|----------|----------------|
| 25-34    | -9657    | 766            |
| 45-54    | 1692     | 737            |
| 55-64    | -7878    | 827            |
| 65-69    | -27324   | 1248           |
| LT HS    | -7989    | 965            |
| Some PS  | 3813     | 738            |
| BA/BSc   | 5838     | 1157           |
| Post BA  | 17300    | 736            |

# Part of a Regression

Regression are cool because you can look at **part** of a regression.

Consider the part with the immigrant cohort effects.

Holding age and education constant, if you compare a Canadian-born person with a Temporary resident, the latter earns \$20,328 less.

Immigrants from the 1960s earn only \$1538 less, and the confidence interval for that is about [-1500, 4500]: it could be zero dollars less!

| Variable          | Est    | SE   |
|-------------------|--------|------|
| Cdn-born 35-44 HS | 37629  | 742  |
| Before 1950       | -3662  | 5999 |
| 1950s             | 952    | 2101 |
| 1960s             | -1538  | 1446 |
| 1970s             | -4382  | 1020 |
| 1980s             | -10952 | 1057 |
| 1990s             | -18928 | 760  |
| 2000s             | -25217 | 1059 |
| Temporary         | -20328 | 2017 |

# Multivariate Analysis

- With tables of proportions or frequencies (ie. Cross-tabs), you can see concentration.
  - Eg., concentration of immigrants at the low end of the income distribution
- You can also see that they are concentrated in the middle of the age distribution and the top of the education distribution
- Thus, *controlling* for these things is essential.

# Multivariate Analysis

- Regressions allow you to control for lots of things at once.
- One-way analysis on the basis of age might tell you one thing (eg., earnings seem okay)
- One-way analysis on the basis of education might tell you another (eg., earnings are low)
- Multivariate analysis can resolve this by controlling for both at the same time.

# R-squared

- The regression model has just three things: year of arrival, age and education.
- These three things 'explain' 7.5% of the variation in earnings from wages and salaries.
  - Explain, as in *are correlated with*
  - Not explain, as in *causes*
- What explains the other 92.5%?
  - Things that explain earnings that are correlated with arrival, age and education
  - Things that explain earnings that are not correlated with arrival, age and education

# Men Earn More---So What

- Men earn more than women. What if we add “male” to the regression?
  - 1970s immigrants have the same sex ratio as Canadian-borns

| Immigrant Cohort | male  |       | Total  |
|------------------|-------|-------|--------|
|                  | 0     | 1     |        |
| Before 1950      | 45.45 | 54.55 | 100.00 |
| 1950s            | 49.57 | 50.43 | 100.00 |
| 1960s            | 52.34 | 47.66 | 100.00 |
| 1970s            | 51.92 | 48.08 | 100.00 |
| 1980s            | 53.08 | 46.92 | 100.00 |
| 1990s            | 54.27 | 45.73 | 100.00 |
| 2000s            | 55.94 | 44.06 | 100.00 |
| Cdn-Born         | 50.10 | 49.90 | 100.00 |
| Temporary        | 55.59 | 44.41 | 100.00 |
| Total            | 51.74 | 48.26 | 100.00 |

# Comparing Regressions

|                   | Coef   | SE   | Coef   | SE   | Coef   | SE    |
|-------------------|--------|------|--------|------|--------|-------|
| Cdn-born 35-44 HS | 37629  | 742  | 29566  | 777  | 48207  | 1360  |
| Before 1950       | -3662  | 5999 | -4387  | 5915 | -5535  | 10462 |
| 1950s             | 952    | 2101 | 778    | 2072 | 1538   | 3807  |
| 1960s             | -1538  | 1446 | -1218  | 1426 | -1345  | 2688  |
| 1970s             | -4382  | 1020 | -4114  | 1006 | -7893  | 1890  |
| 1980s             | -10952 | 1057 | -10523 | 1042 | -15131 | 1975  |
| 1990s             | -18928 | 760  | -18298 | 750  | -25109 | 1433  |
| 2000s             | -25217 | 1059 | -24254 | 1045 | -31833 | 2042  |
| Temporary         | -20328 | 2017 | -19392 | 1989 | -15805 | 3895  |
| 25-34             | -9657  | 766  | -9537  | 755  | -14263 | 1418  |
| 45-54             | 1692   | 737  | 1849   | 726  | 2777   | 1362  |
| 55-64             | -7878  | 827  | -7778  | 815  | -7198  | 1519  |
| 65-69             | -27324 | 1248 | -27016 | 1230 | -35258 | 2312  |
| LT HS             | -7989  | 965  | -8057  | 951  | -11041 | 1789  |
| Some PS           | 3813   | 738  | 3599   | 728  | 3929   | 1366  |
| BA/BSc            | 5838   | 1157 | 6871   | 1142 | 6881   | 2295  |
| Post BA           | 17300  | 736  | 16962  | 726  | 20588  | 1360  |
| male              |        |      | 16172  | 526  |        |       |
|                   |        |      |        |      |        |       |

|                   | Coef   | SE   | Coef   | SE   |
|-------------------|--------|------|--------|------|
| Cdn-born 35-44 HS | 37629  | 742  | 29566  | 777  |
| Before 1950       | -3662  | 5999 | -4387  | 5915 |
| 1950s             | 952    | 2101 | 778    | 2072 |
| 1960s             | -1538  | 1446 | -1218  | 1426 |
| 1970s             | -4382  | 1020 | -4114  | 1006 |
| 1980s             | -10952 | 1057 | -10523 | 1042 |
| 1990s             | -18928 | 760  | -18298 | 750  |
| 2000s             | -25217 | 1059 | -24254 | 1045 |
| Temporary         | -20328 | 2017 | -19392 | 1989 |

- Check out just the top left panel: on the left there's a regression without *male*, on the right a regression with *male*.
- Adding *male* doesn't do anything
- This is because male, though important to earnings, is not very correlated with immigrant arrival cohort.

# Just Men's Earnings

This is the rightmost panel: a regression with just men in it.

The same pattern emerges: immigrants from earlier cohorts have higher earnings, and immigrants from recent cohorts have very low earnings.

|                   | Coef   | SE    |
|-------------------|--------|-------|
| Cdn-born 35-44 HS | 48207  | 1360  |
| Before 1950       | -5535  | 10462 |
| 1950s             | 1538   | 3807  |
| 1960s             | -1345  | 2688  |
| 1970s             | -7893  | 1890  |
| 1980s             | -15131 | 1975  |
| 1990s             | -25109 | 1433  |
| 2000s             | -31833 | 2042  |
| Temporary         | -15805 | 3895  |
| 25-34             | -14263 | 1418  |
| 45-54             | 2777   | 1362  |
| 55-64             | -7198  | 1519  |
| 65-69             | -35258 | 2312  |
| LT HS             | -11041 | 1789  |
| Some PS           | 3929   | 1366  |
| BA/BSc            | 6881   | 2295  |
| Post BA           | 20588  | 1360  |
| male              |        |       |
| just men          | yes    |       |

# Endogeneity

- Adding Male to the regression changed nothing important (because it wasn't correlated with cohort)
  - But let's stick with just men.
- What about visible minority status? It is correlated.

| Immigrant Cohort | VM    |       | Total  |
|------------------|-------|-------|--------|
|                  | 0     | 1     |        |
| Before 1950      | 94.44 | 5.56  | 100.00 |
| 1950s            | 86.21 | 13.79 | 100.00 |
| 1960s            | 69.38 | 30.62 | 100.00 |
| 1970s            | 35.74 | 64.26 | 100.00 |
| 1980s            | 18.34 | 81.66 | 100.00 |
| 1990s            | 14.61 | 85.39 | 100.00 |
| 2000s            | 17.57 | 82.43 | 100.00 |
| Cdn-Born         | 90.71 | 9.29  | 100.00 |
| Temporary        | 37.78 | 62.22 | 100.00 |
| Total            | 62.70 | 37.30 | 100.00 |

|                         | Coef   | SE    | Coef   | SE    |
|-------------------------|--------|-------|--------|-------|
| White Cdn-born 35-44 HS | 48207  | 1360  | 49041  | 1364  |
| Before 1950             | -5535  | 10462 | -5361  | 10448 |
| 1950s                   | 1538   | 3807  | 2343   | 3804  |
| 1960s                   | -1345  | 2688  | 998    | 2708  |
| 1970s                   | -7893  | 1890  | -2463  | 2060  |
| 1980s                   | -15131 | 1975  | -8222  | 2233  |
| 1990s                   | -25109 | 1433  | -17912 | 1799  |
| 2000s                   | -31833 | 2042  | -25021 | 2286  |
| Temporary               | -15805 | 3895  | -11181 | 3952  |
| 25-34                   | -14263 | 1418  | -13452 | 1421  |
| 45-54                   | 2777   | 1362  | 2625   | 1360  |
| 55-64                   | -7198  | 1519  | -7576  | 1519  |
| 65-69                   | -35258 | 2312  | -35378 | 2309  |
| LT HS                   | -11041 | 1789  | -10752 | 1787  |
| Some PS                 | 3929   | 1366  | 3540   | 1365  |
| BA/BSc                  | 6881   | 2295  | 7352   | 2293  |
| Post BA                 | 20588  | 1360  | 20674  | 1358  |
| vm                      |        |       | -9523  | 1444  |
|                         |        |       |        |       |

# Endogeneity

- Those numbers changed a lot!
- Visible minority status is correlated with time of arrival: immigrants in later cohorts were more likely to be visible minorities.
- Visible minorities have lower earnings.
- Thus, if you leave visible minority out of the regression, you think you are measuring a cohort effect, but it is really a visible minority effect.
- This is a form of *endogeneity*. If a coefficient suffers from endogeneity, then its estimated value is not informative of the effect of the variable of interest.

# Correlation and Causation

- Two things make correlation different from causation.
  - *missing variables*: things (eg, earnings and arrival cohort) may be correlated, but that correlation may be driven by left-out or unobserved stuff (eg, country-of-birth).
  - *reverse causation*: things can be correlated, but you don't know which causes which.
    - Eg, if Chinese-born people fare well in B.C., then they may choose to move here. This means that their high wages *cause* their clustering here.

# Causation and Endogeneity

- If two variables are correlated but are not causally related, the econometric problem is one of endogeneity (just as above).

# Policy Needs Isolated Effects

- If we want to know *whether or not* recent cohorts fare badly, a table will do fine.
- If we want to know *why* they fare badly, we need to control out stuff
  - Eg., part of their ill-fortune is driven by their ethnic origin, not when they came.

# Let's Educate Immigrants

Education pays off no matter how you slice it, controlling for age or not.

Should we invest in educating immigrants as a matter of policy?

|          | Coef   | SE   | Coef   | SE   |
|----------|--------|------|--------|------|
| LT HS    | -11041 | 1789 | -10752 | 1787 |
| Some PS  | 3929   | 1366 | 3540   | 1365 |
| BA/BSc   | 6881   | 2295 | 7352   | 2293 |
| Post BA  | 20588  | 1360 | 20674  | 1358 |
| vm       |        |      | yes    |      |
| just men | yes    |      | yes    |      |

# Policy Needs Causal Effects

- Does the table tell what happens if we get more educated immigrants?
- A lot of missing variables here:
  - Quality of education; relevance of education; place of education.
  - Are people who get educated smarter than people who don't? If so, then educating a given person won't raise their wage as much as the difference in earnings might suggest.

# Instrumental Variables

- Education is endogenous--that is we don't really believe that we have included all the correlated missing variables.
- We could use an *instrument* to solve this problem.
  - An *instrument* is a variable the is correlated with the problem variable (education in this case) but not with all those correlated missing variables
  - Bit too tricky to define right now.

# Experimental Approaches

- The idea of experiments, in the lab, in the field, or natural experiments, is:
  - hold everything constant while changing something of interest
  - Measure the difference across its values
  - Since nothing else is changing, there are no missing variables.
  - Since the experimenter is changing the values of the variable, there is no reverse causation.

# Natural Experiments

- Lab experiments don't work so well for economic questions, eg, do immigrant arrivals affect native-born wages? Need a big lab.
- A *natural* experiment addresses reverse causality: use a change that for sure is not driven by your supposed dependent variable.
  - Eg, the Mariel Boatlift put more than 100,000 Cubans into Florida.
  - This was not due to high wages in Florida; it was due to Castro emptying prisons.
  - It had no effect on native-born white or hispanic wages (David Card 1990)

# Field Experiments

- Often, you don't get the natural experiment you wish for.
- A field experiment uses data created by the investigator in such a way that there is little or no missing variables problem.
- Phil Oreopoulos (2009) (in MBC-supported research) sent out 8000 fake resumes to figure out whether and why immigrants have trouble finding work.

# Oreopoulous' Study

- Field experiment with 8,000 mock resumes, emailed in response to newspaper and online job postings
- Randomize name, experience, education, and other characteristics, send, and compare callback rate differences
- Offers clear and convincing evidence on why recent immigrants fare poorly in the Canadian labour market

# Example: Different experience

2390 Credit Valley Road, Mississauga, ON, L5M 4E6, (905) 901-3811, tara.singh45@yahoo.ca

## Tara Singh

### Professional Summary

- Motivated professional with demonstrated analytical abilities and investment research skills.
- Highly-developed planning and analytical skills from 5 years of relevant experience.
- Reliable, dependable, self motivated, flexible and efficient .
- Outstanding knowledge of Microsoft Office.
- Able to work extremely well under pressure.

### Experience

KPMG Inc. 2006 - 2007 Mumbai, India

#### Senior Accountant

- Full cycle of accounting including payroll, accounts payable/receivable, account reconciliations and period end closing, reported to CFO Formulated and graphed monthly quarterly sales analysis spreadsheets to re-focus sales activity and achieve a 10% gain in sales. Designed daily cash flow report summarizing inflows and outflows to numerous bank accounts resulting in 5% saving in cost of funds. Drafted the GL procedure manual, automated the month end reporting process instead of manual, designed and analyzed an efficient spreadsheet for management report.

KPMG Inc. 2005 - 2006 Mumbai, India

#### Accounting Supervisor

- Supervised 10 A/P and A/R; followed up payment to projects and maintained daily accounts payable system; Reorganized A/R system Maintained daily transaction records, processed invoices, tracked expenses, filing etc. Reconciled bank statements, directed all cash activities, prepared tax documents and annual financial statements.

Blue Star Infotech Ltd. 2004 - 2005 Mumbai, India

#### Accountant

- Coordinated distribution of invoices and classified transactions. Augmented the A/P and A/R process, posted to A/R and A/P journals, prepared drafting P&L statements and monthly balance sheet consolidation.

### Education

Indian Institute of Management 2000 - 2004 Bangalore, India

- Bachelor of Science, Economics

### Additional Interests and Activities

#### World Traveller

- Travelled to thirty-one countries on five continents.

#### Big Sisters

- Mentor for disadvantaged youth.

#### Other Activities

- Competitive squash player, classical piano player, recreational photographer.

2390 Credit Valley Road, Mississauga, ON, L5M 4E6, (905) 901-3811, tara.singh45@yahoo.ca

## Tara Singh

### Professional Summary

- Motivated professional with demonstrated analytical abilities and investment research skills.
- Highly-developed planning and analytical skills from 5 years of relevant experience.
- Reliable, dependable, self motivated, flexible and efficient .
- Outstanding knowledge of Microsoft Office.
- Able to work extremely well under pressure.

### Experience

KPMG Inc. 2006 - 2007 Toronto, Canada

#### Senior Accountant

- Full cycle of accounting including payroll, accounts payable/receivable, account reconciliations and period end closing, reported to CFO Formulated and graphed monthly quarterly sales analysis spreadsheets to re-focus sales activity and achieve a 10% gain in sales. Designed daily cash flow report summarizing inflows and outflows to numerous bank accounts resulting in 5% saving in cost of funds. Drafted the GL procedure manual, automated the month end reporting process instead of manual, designed and analyzed an efficient spreadsheet for management report.

KPMG Inc. 2005 - 2006 Toronto, Canada

#### Accounting Supervisor

- Supervised 10 A/P and A/R; followed up payment to projects and maintained daily accounts payable system; Reorganized A/R system Maintained daily transaction records, processed invoices, tracked expenses, filing etc. Reconciled bank statements, directed all cash activities, prepared tax documents and annual financial statements.

Mackenzie Financial Corporation 2004 - 2005 Toronto, Canada

#### Accountant

- Coordinated distribution of invoices and classified transactions. Augmented the A/P and A/R process, posted to A/R and A/P journals, prepared drafting P&L statements and monthly balance sheet consolidation.

### Education

Indian Institute of Management 2000 - 2004 Bangalore, India

- Bachelor of Science, Economics

### Additional Interests and Activities

#### World Traveller

- Travelled to thirty-one countries on five continents.

#### Big Sisters

- Mentor for disadvantaged youth.

#### Other Activities

- Competitive squash player, classical piano player, recreational photographer.

# Example: Different name

5354 Russell View Road, Mississauga, ON, L5M 5V8, (647) 477-2873, martin.john15@gmail.com

## John Martin

### Professional Summary

- Experienced in various business aspects; accounting, customer relations, computer training, sales, marketing, negotiations, presentations, and office operations.
- Highly motivated.
- Experienced multi-tasker.
- Analytical and detail-oriented, problem solver.
- Efficient with time management skills.
- Excellent skills in Visual Basic, HTML and Microsoft Office.

### Experience

KPMG Corp. 2006 to date Toronto, Ontario

#### Financial Analyst

- Assisted the Corporate Finance Director with budgeting and forecasting exercises. Identified, explained and communicated variances for operating plans and latest forecast. Examined the feasibility of business projects and prepare a plan of action based on financial analysis. Reconciled monthly bank statements entries via AS400.

ZAC Marketing Inc 2004 - 2006 Toronto, Ontario

#### Actuarial Analyst

- Performed actuarial and statistics analysis of risk to provide the underwriting department with keys contract valuation metrics. Developed actuarial models used for pricing and/or risk management. Performed segmentation analysis on the behalf of insurance companies to determine best and worst performing products/classes and recommend strategies for growing/correcting those areas as appropriate. Examined expert risk reports on larges individual corporate risks. Improved the decision making process significantly and the quality of internal statistical and technical reporting documents by creating an Access based program that offered a wide range analyses of the company's portfolio of reinsurance contracts. Improved the average technical account reconciliation time by more than fifty percent.

FGF Brands Inc. 2002 - 2004 Toronto, Ontario

#### Investment Analyst

- Independently performed fundamental research on assigned securities (distribution sectors). Participated in the decision making process with respect to portfolio management by making buy, sell and hold recommendations. Analyzed and tracked key data and statistics related to individual stocks and portfolios. Created and maintained financial models for stock and portfolio analysis. Produced performance reports that include analyses of returns, risk, added value and portfolio characteristics. Created an Excel-based application that calculates more than 20 financial and operational ratios. Elected employee of the month four times.

### Education

University of Waterloo 1998 - 2002 Waterloo, Ontario

- Bachelor of Commerce, Accounting

### Additional Interests and Activities

**World Traveller:** Travelled to thirty-one countries on five continents.

**Big Brothers:** Mentor for disadvantaged youth.

**Other Activities:** Competitive squash player, classical piano player, recreational photographer.

5354 Russell View Road, Mississauga, ON, L5M 5V8, (647) 477-2873, zhang.long11@gmail.com

## Zhang Long

### Professional Summary

- Experienced in various business aspects; accounting, customer relations, computer training, sales, marketing, negotiations, presentations, and office operations.
- Highly motivated.
- Experienced multi-tasker.
- Analytical and detail-oriented, problem solver.
- Efficient with time management skills.
- Excellent skills in Visual Basic, HTML and Microsoft Office.

### Experience

KPMG Corp. 2006 to date Toronto, Ontario

#### Financial Analyst

- Assisted the Corporate Finance Director with budgeting and forecasting exercises. Identified, explained and communicated variances for operating plans and latest forecast. Examined the feasibility of business projects and prepare a plan of action based on financial analysis. Reconciled monthly bank statements entries via AS400.

ZAC Marketing Inc 2004 - 2006 Toronto, Ontario

#### Actuarial Analyst

- Performed actuarial and statistics analysis of risk to provide the underwriting department with keys contract valuation metrics. Developed actuarial models used for pricing and/or risk management. Performed segmentation analysis on the behalf of insurance companies to determine best and worst performing products/classes and recommend strategies for growing/correcting those areas as appropriate. Examined expert risk reports on larges individual corporate risks. Improved the decision making process significantly and the quality of internal statistical and technical reporting documents by creating an Access based program that offered a wide range analyses of the company's portfolio of reinsurance contracts. Improved the average technical account reconciliation time by more than fifty percent.

FGF Brands Inc. 2002 - 2004 Toronto, Ontario

#### Investment Analyst

- Independently performed fundamental research on assigned securities (distribution sectors). Participated in the decision making process with respect to portfolio management by making buy, sell and hold recommendations. Analyzed and tracked key data and statistics related to individual stocks and portfolios. Created and maintained financial models for stock and portfolio analysis. Produced performance reports that include analyses of returns, risk, added value and portfolio characteristics. Created an Excel-based application that calculates more than 20 financial and operational ratios. Elected employee of the month four times.

### Education

University of Waterloo 1998 - 2002 Waterloo, Ontario

- Bachelor of Commerce, Accounting

### Additional Interests and Activities

**World Traveller:** Travelled to thirty-one countries on five continents.

**Big Brothers:** Mentor for disadvantaged youth.

**Other Activities:** Competitive squash player, classical piano player, recreational photographer.

# Field Experiment Setup

- Respond only to internet or newspaper postings (e.g. Workopolis, Job Bank, Craigslist,...) May – Nov. 2008
- 4 resumes per posting, sent over 2-3 days:
- 0: Can. name, Can. education, Can. experience
- 1: foreign name, Can. education, Can. Exp.
- 2: foreign name, foreign education, Can. Exp.
- 3: 4: foreign name, foreign educ., some or all foreign exp., respectively

### Callback Rates by Resume Type

(Difference Compared to Type 0)

[Standard Error of Difference, \* indicates sign. Diff. compared to prev. type]

{Callback Ratio: Type 0 / Type}

|        |                                             | Ethnic Origin  |                                           |                                           |                                           |                                        |                                           |
|--------|---------------------------------------------|----------------|-------------------------------------------|-------------------------------------------|-------------------------------------------|----------------------------------------|-------------------------------------------|
|        |                                             | English-Canada | India                                     | China                                     | Pakistan                                  | Britain                                | India/China/Pakistan                      |
| Type 0 | English Name<br>Cdn Educ/Exp                | 0.158          |                                           |                                           |                                           |                                        |                                           |
| Type 1 | Foreign Name<br>Cdn Educ<br>Cdn Exp         |                | 0.121<br>(-0.037)<br>[0.019]*<br>{1.31}   | 0.108<br>(-0.050)<br>[0.018]***<br>{1.46} | 0.11<br>(-0.048)<br>[0.016]***<br>{1.44}  | NA                                     | 0.113<br>(-0.045)<br>[0.011]***<br>{1.40} |
| Type 2 | Foreign Name<br>Foreign Educ<br>Cdn Exp     |                | 0.122<br>(-0.036)<br>[0.022]<br>{1.30}    | 0.094<br>(-0.064)<br>[0.020]<br>{1.68}    | 0.14<br>(-0.018)<br>[0.027]<br>{1.13}     | 0.129<br>(-0.029)<br>[0.019]<br>{1.22} | 0.114<br>(-0.044)<br>[0.014]<br>{1.39}    |
| Type 3 | Foreign Name<br>Foreign Educ<br>Mixed Exp   |                | 0.075<br>(-0.083)<br>[0.019]***<br>{2.11} | 0.103<br>(-0.055)<br>[0.021]<br>{1.53}    | 0.078<br>(-0.080)<br>[0.020]***<br>{2.03} | 0.157<br>(-0.001)<br>[0.023]<br>{1.01} | 0.088<br>(-0.070)<br>[0.013]***<br>{1.80} |
| Type 4 | Foreign Name<br>Foreign Educ<br>Foreign Exp |                | 0.051<br>(-0.107)<br>[0.017]**<br>{3.10}  | 0.053<br>(-0.105)<br>[0.018]***<br>{2.98} | 0.052<br>(-0.106)<br>[0.015]**<br>{3.04}  | 0.141<br>(-0.017)<br>[0.021]<br>{1.12} | 0.052<br>(-0.106)<br>[0.011]***<br>{3.04} |

# Oreopoulous' Results

- Name discrimination likely illegal
  - “Subsection 11 (1) of the Code also establishes that the right of a person under Part I is infringed where a requirement, qualification or factor exists that is not a prohibited ground of discrimination, but that results in the exclusion or restriction of a group of persons who are identified by a prohibited ground of discrimination, unless the requirement or factor is reasonable and genuine in the circumstances, subject to hardship on the employer”
- For those with 4-6 years Can. Experience, minimal differences for foreign vs. Can education
- Lower returns to foreign experience not because of type of job or employer
- Brits do fine

# Food For Thought

- One-way tables are not enough:
  - Missing variables may be correlated, and may in fact be driving things
  - Multivariate analysis can fix up missing variables problems, and allow you to get at causation
- Statistical precision matters:
  - if things look different, but are not statistically distinguished (eg, because they are only 1 std err apart), then they may not be different at all.
- Regression analysis is a multivariate tool that lets you think about each variable separately.
- Natural experiments, field experiments and instrument variables help with both missing variables and reverse causation, and thus allow you to get at causation.