The Classical Model

Multicollinearity and Endogeneity

What Information Does OLS use?

- Kennedy (1981, 2002 and 2003) presents Ballentine Diagrams
  - these show how OLS uses information to estimate regression coefficients
  - and, where the variance (precision) of those estimated coefficients comes from
- See his 2002 paper on the web: http://www.amstat.org/publications/jse/v10n1/kennedy.html

OLS Ballentine

- Where’s Cov(X,Y)?
- What is its value?
- Where’s epsilon?

Ballentines with 2 Regressors

- size of circles gives variance.
- overlap gives identification of estimator.
- size of overlap gives variance of estimator.
- where's the error term?
- 3 possible estimators of coefficients:
  - (x1,y) overlap, (x2,y) overlap (use x1,x2 overlap twice;
  - divide the (x1,x2) overlap
  - disgard the (x1,x2) overlap
- What happens if x1,x2 have higher covariance?

Leaving Out the Constant

- Compare the formula for the slope with a constant:

$$\hat{\beta}_1 = \frac{\sum_i \left(X_i - \bar{X} \right)\left(Y_i - \bar{Y} \right)}{\sum_i \left(X_i - \bar{X} \right)^2}$$

- To that without a constant

$$\hat{\beta}_1 = \frac{\sum_i \left(X_i Y_i \right)}{\sum_i \left(X_i \right)^2}$$

Violating Assumption 6: $Cov(X_1, X_2) \neq 0$

- Recall we assume that no independent variable is a **perfect linear function** of any other independent variable.
  - If a variable X_1 can be written as a perfect linear function of X_2 , X_3 , etc., then we say these variables are **perfectly collinear**.
  - When this is true of more than one independent variable, they are **perfectly multicollinear.**
- Perfect multicollinearity presents technical problems for computing the least squares estimates.
  - Example: suppose we want to estimate the regression: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ where $X_1 = 2X_2 + 5$. That is, X_1 and X_2 are perfectly collinear. Whenever X_2 increases by one unit, we see X_1 increase by 2 units, and Y increase by $2\beta_1 + \beta_2$ units. It is completely arbitrary whether we attribute this increase in Y to X_1, to X_2, or to some combination of them. If X_1 is in the model, then X_2 is completely redundant: it contains **exactly** the same information as X_1 (if we know the value of X_1, we know the value of X_2 **exactly**, and vice versa). Because of this, there is no unique solution to the least squares minimization problem. Rather, there are an infinite number of solutions.
  - Another way to think about this example: β_1 measures the effect of X_1 on Y, holding X_2 constant. Because X_1 and X_2 always vary (exactly) together, there's no way to estimate this.

Imperfect Multicollinearity

- It is quite rare that two independent variables have an **exact** linear relationship
  - it's usually obvious when it does happen: e.g., the "dummy variable trap"
- However it is very common in economic data that two (or more) independent variables are strongly, but not exactly, related
  - in economic data, everything affects everything else
- Example:
  - perfect collinearity: $X_{1i} = \alpha_0 + \alpha_1 X_{2i}$
  - imperfect collinearity: $X_{1i} = \alpha_0 + \alpha_1 X_{2i} + \zeta_i$ where ζ_i is a stochastic error term
- Examples of economic variables that are strongly (but not exactly) related:
  - income, savings, and wealth
  - firm size (employment), capital stock, and revenues
  - unemployment rate, exchange rate, interest rate, bank deposits
- Thankfully, economic theory (and common sense!) tell us these variables will be strongly related, so we shouldn't be surprised to find that they are ...
- But when in doubt, we can look at the **sample correlation** between independent variables to detect imperfect multicollinearity
- When the sample correlation is big enough, Assumption 6 is "almost" violated

Consequences of Multicollinearity

- Least squares estimates are still unbiased
- recall that only Assumptions 1-3 of the CLRM (correct specification, zero expected error, exogenous independent variables) are required for the least squares estimator to be unbiased
- since none of those assumptions are violated, the least squares estimator remains unbiased

Consequences of Multicollinearity

- The least squares estimates will have big standard errors
- this is the main problem with multicollinearity
- we're trying to estimate the marginal effect of an independent variable holding the other independent variables constant.
- But the strong linear relationship among the independent variables makes this difficult – we always see them move together
- That is, there is very little information in the data about the thing we're trying to estimate
- Consequently, we can't estimate it very precisely: the standard errors are large

Consequences of Multicollinearity

- The computed t-scores will be small
- think about what 1 & 2 imply for the sampling distribution of the least squares estimates:
- the sampling distribution is centered over the true parameter values (the estimates are unbiased) but the sampling variance is very large (draw a picture)
- thus we're "likely" to obtain estimates that are "far" from the true parameter values
- because the estimates are very imprecise, we have a difficult time rejecting any null hypothesis, because we know that even when the null is true, we could find an estimate far from the hypothesized value
- from the formula for the t statistic:

$$t = \frac{\hat{\beta}_k - \beta_{0k}}{se(\hat{\beta}_k)}$$

  we see that because the standard error is big, the t statistic is small (all else equal), and thus we are unlikely to reject H0

Consequences of Multicollinearity

- Small changes in data or specification lead to large changes in parameter estimates
- adding/deleting an independent variable, or adding/deleting a few observations can have huge effects on the parameter estimates
- why?
- think about the sampling distribution of the least squares estimates: it's very spread out around the true coefficient values. In different samples/specifications we are likely to get very different estimates
- because there is so little independent variation in the independent variables, the least squares estimator puts a lot of weight on small differences between them. Small changes in sample/specification that affect these small differences (even a little bit) get a lot of weight in the estimates
- e.g., if two variables are almost identical in most of the sample, the estimator relies on a few observations where they move differently to distinguish between them. Dropping one of these observations can have a big effect on the estimates.

Detecting Multicollinearity

- It's important to keep in mind that most economic variables are correlated to some degree
  - that is, we face **some** multicollinearity in **every** regression that we run
- The question is how much? And is it a problem?
- We've seen one method of detecting collinearity already: look at the sample correlation between independent variables.
  - rule of thumb: sample correlation > 0.8 is evidence of severe collinearity
  - problem: if the collinear relationship involves more than 2 independent variables, you may not detect it this way
- Look at Variance Inflation Factors (VIFs)
  - regress each independent variable X_{ji} on all the other independent variables
  - $$X_{ji} = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_{j-1} X_{j-1,i} + \alpha_{j+1} X_{j+1,i} + \alpha_k X_{ki} + \varepsilon_i$$
  - collect the R^2 of this regression (call it R_j^2), and compute the VIF: $\text{VIF}(X_j) = 1 / (1 - R_j^2)$
  - Rule of thumb: $\text{VIF}(X_j) > 5$ is evidence of severe multicollinearity

Remedies for Multicollinearity

- Get more data
  - this is always a good idea, and is the best remedy for multicollinearity when it is possible
  - basically, the multicollinearity problem is just that there's not enough independent variation in the data to separately identify marginal effects. If we get more data, we got more variation to identify these effects.
- Do nothing
  - we know our estimates are unbiased
  - put up with the inflated standard errors
- The text has a couple of other suggestions ...
  - drop an irrelevant variable – why is it there in the first place? beware omitted variable bias!
  - transform the multicollinear variables – only if it's consistent with economic theory & common sense

Violating Assumption 3: $Cov(X_i, \varepsilon_i) = 0$

- We saw that correlated missing regressors induce bias.
- So does biased sample selection and reverse causality.
- Consider correlated missing regressors.

Omitted Variables (review)

- Suppose the **true DGP** is:

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

  but we incorrectly estimate the regression model:

  $$Y_i = \beta_0^* + \beta_1^* X_{1i} + \varepsilon_i^*$$

  – *example: Y is earnings, X_1 is education, and X_2 is "work ethic" – we don't observe a person's work ethic in the data, so we can't include it in the regression model*
- That is, we **omit** the variable X_2 from our model
- What is the consequence of this?
- Does it mess up our estimates of β_0 and β_1?
  – it definitely messes up our **interpretation** of β_1. With X_2 in the model, β_1 measures the marginal effect of X_1 on Y **holding X_2 constant**. We can't hold X_2 constant if it's not in the model.
  – Our estimated regression coefficients may be **biased**
  – The estimated β_1 thus measures the marginal effect of X_1 on Y **without holding X_2 constant**. Since X_2 is in the error term, the error term will covary with X_1 if X_2 covaries with X_1 .

Omitted Variables (Review)

$$E\left[\hat{\beta}_1\right] = E\left[\frac{\sum_i \left(X_{1i} - \bar{X}_1\right)\left(Y_i - \bar{Y}\right)}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2}\right] = E\left[\frac{\sum_i \left(X_{1i} - \bar{X}_1\right)\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i - \beta_0 - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \bar{\varepsilon}\right)}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2}\right]$$

$$= E\left[\frac{\sum_i \left(X_{1i} - \bar{X}_1\right)\left(\beta_1\left(X_{1i} - \bar{X}_1\right) + \beta_2\left(X_{2i} - \bar{X}_2\right) + \varepsilon_i - \bar{\varepsilon}\right)}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2}\right]$$

$$= E\left[\frac{\beta_1 \sum_i \left(X_{1i} - \bar{X}_1\right)^2}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2} + \frac{\sum_i \left(X_{1i} - \bar{X}\right)\left(\beta_2\left(X_{2i} - \bar{X}_2\right) + \varepsilon_i - \bar{\varepsilon}\right)}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2}\right]$$

$$= \beta_1 + \beta_2 E\left[\frac{\sum_i \left(X_{1i} - \bar{X}\right)\left(\left(X_{2i} - \bar{X}_2\right) + \varepsilon_i - \bar{\varepsilon}\right)}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2}\right] = \beta_1 + \frac{\beta_2}{\sum_i \left(X_{1i} - \bar{X}_1\right)^2} E\left[\sum_i \left(X_{1i} - \bar{X}\right)\left(X_{2i} - \bar{X}_2\right)\right]$$

$$= \beta_1 + \frac{\beta_2}{\sum_i \left(X_{1i} - \bar{X}\ \right)^2} nCov\left[X_1, X_2\right] = \beta_1 + \beta_2 \frac{Cov\left[X_1, X_2\right]}{Var\left[X_1\right]}$$

The estimated parameter is **biased**, with bias linear in the true parameter on the left-out variable, and the covariance of the left-out variable with the included variable.

General Endogeneity Bias

- In general, we say that a variable *X* is endogenous if it is correlated with the model error term. Endogeneity always induces bias.
- In the 1 included regressor case, we always have

$$E\left[\hat{\beta}_1\right] = \beta_1 + E\left[\frac{\sum_i \left(X_i - \bar{X}\right)\left(\varepsilon_i - \bar{\varepsilon}\right)}{\sum_i \left(X_i - \bar{X}\right)^2}\right]$$

$$= \beta_1 + \frac{Cov\left[X_1, \varepsilon\right]}{Var\left[X_1\right]}$$

Endogeneity in a Scatter-Plot

- Endogeneity is easy to draw.
- Consider a 1 variable (with intercept) model
- Let the conditional mean of the error term also rise linearly with the included variable
- Draw the true regression line and the data
- The OLS regression line will pick up both the slope of Y in X and the slope of the conditional mean of the error with respect to X.

General Endogeneity Bias

- Endogeneity bias shows up in the Ballentine diagrams.
- Correlated missing regressor: x2 is invisible.
- What does the OLS estimator do to the (x1,x2) overlap?
- more generally, some of what looks like (x1,y) is really the model error term, and not (x1,y).

Causes of Endogeneity

- Correlated Missing Regressors
- Sample selection
  - what if you select the sample on the basis of something correlated with epsilon?
  - For example, if you only look at workers, and exclude non-workers, the latter might have different epsilons from the former.
- Reverse causality
  - what if *Y* causes *X*? Since *Y* contains epsilon, then variation in epsilon would show up in *X*.

Endogeneity in Practise

- Is there a correlated missing regressor problem in the immigration cohort analysis you did in assignment 2?
- What is the correlation that might be a problem? What direction would the bias go?
- Is there a sample selection problem in the paper you read for assignment 3?
- What is the correlation that might be a problem? What direction would the bias go?

Correcting for Endogeneity

- Correlated Missing Regressors
  - Include the missing data
- Sample Selection
  - Narrow the generality of interpretation: no longer applies to the population, but rather to the chunk of the population satisfying sample restrictions
- Reverse Causality is tougher

Correcting for Endogeneity

- Endogeneity is like pollution in the *X*.
- You need information that allows you to pull the pollution out of the *X*.
- Including missing regressors is like identifying the pollution exactly, so that you can just use the *X* that is uncorrelated with that pollution.
- Alternatively, you could find a part of the variation in *X* that is unpolluted by construction.

Instrumental Variables

- *Instruments* are variables, denoted *Z*, that are correlated with *X*, but uncorrelated with the model error term by assumption or by construction.
- Cov(Z,e)=0, so in the Ballentine, Z and the error term have no overlap.
- But, (Z,X) do overlap

2-Stage Least Squares

- Regress *X* on *Z*
  - generate $\hat{X}$ =*E[X|Z]*, the predicted value of *X* given *Z*.
  - This is "clean". Since *Z* is uncorrelated with the model error term, so is any linear function of *Z*.
- Regress *Y* on $\hat{X}$
- This regression does not suffer from endogeneity
- But it does suffer from having less variance in its regressor.

Good Instruments

- Instrumental variables just push the problem backwards.
  - Is your instrument actually exogenous? That is, is it really uncorrelated with the model error term?
- Consider regressing wages on education level
  - education is endogenous: unobserved ability is correlated with both education and wages
  - instrument would have to be correlated with education level of a person, but not their ability level
    - compulsory schooling age (Angrist and Krueger); distance to college (Card); presence of an ag-grant college (Moretti); etc.

Experiments

- With correlated missing regressors, you can undo the endogeneity problem *completely* if you are sure that your variable of interest if randomly distributed with respect to all missing regressors
  - Ballentine with 2 disjoint regressors shows this.
- experimental methods enforce this randomness by construction

Experiments, example

- In assignment 3, you read a paper that showed that visible minority men have lower earnings than white men.
- But, is the regressor (visible minority) exogenous?
  - correlated missing regressors that affect earnings could include field-of-study, age, years of schooling, grades, etc

A Field Experiment

- Oreopoulous (2010) conducts a field experiment to try to figure out whether results like mine are driven by endogeneity
- The whole point is that he can randomize the "treatment" (in this case, having a minority name).
  - If it is random, then it has no covariance with any missing regressors

What are resume audit studies?

- Field experiments to test how characteristics on resume affect decisions to contact applicant for job interview
  - Similarities and differences between applicants are perfectly known
  - Relatively cheap, so sample can be large
- Bertrand and Mullainathan (2004)
  - Resumes with randomly assigned white sounding names receive 1.5 times more callbacks than resumes with black sounding names
- Other resume audit studies:
  - Swedish names versus middle-eastern names in Sweden
  - Old versus young (by date of HS graduation)
  - Mother/Father versus single (by extracurricular activities)
- This one:
  - First audit study to look at immigrants that differ from natives by potentially more than one factor
  - First resume audit study in Canada

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.

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

Random Assignment

- What if some firms are racist and others are not?
  - Because the treatment (getting a minority resume) is randomly assigned, the estimated effect is the *average* of the racist and nonracist firm behaviour.
  - You could check how the treatment effect (regression estimate) varies with observable firm characteristics.
    - E.g., Oreopoulous finds bigger effects in smaller firms.

Random Assignment

- *Random assignment* is enough to undo the effects of correlated missing regressor induced endogeneity.
- But, it is too much. Angrist and Pischke (2008) Mostly Harmless Econometrics have a nice exposition of the fact that all you need is *conditionally random assignment*.
- That is, if the regressor-of-interest is independent of the missing regressors, *conditional on the included regressors*, then the OLS coefficient estimate is unbiased.

Conditional Independence

- Suppose that you have data on whether patients receive a certain treatment, and on their health outcome.
- Suppose that
  - The treatment effect depends on patient health
  - Treatments are offered to unhealthy patients
  - Healthiness is completely explained by age
- Then, regressing health outcome on treatment yields endogeneous results, but regressing health outcome on treatment and age does not.

Conditional Independence

- Demanding conditional independence feels a lot like asking that all missing regressors be included.
- But, it is less than asking for independence.
- When designing data, it can be useful:
  - In the previous example, we don't need hospitals to assign treatments randomly (which would be unethical if people need the treatment).
  - We need to assign treatments randomly *conditional on the other observables*.