March Family Internship Fund

- The Economics Department would like to remind you about the March Family Internship Fund
- It is a scholarship for econ majors so that they can afford to do an unpaid internship
- Application deadline is May 2nd
- Additional info: www.econ.ucdavis.edu/undergraduates_internship_info.cfm?id=1631
- Application website: www.econ.ucdavis.edu/application/app.cfm

Final Exam Details

- The final is Thursday, March 17 from 10:30am to 12:30pm in the regular lecture room
- The final is cumulative (multiple choice will be a roughly 50/50 split between material since the second midterm and old material, short answer will be focused on the new material)
- The old finals are a good guide to the format and length of the exam as well as the division of the exam between old and new material
- The formula sheet will be posted tomorrow on Smartsite
- Office hours during exam week: Monday 2pm-4pm, Tuesday 10am-12pm, Wednesday 10am-12pm

Review: Model Misspecification Problems

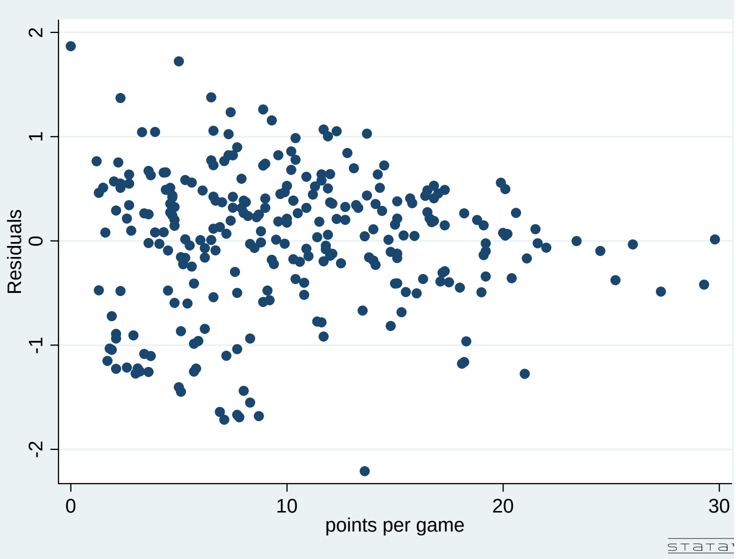
Some of the issues we've covered so far:

- Omitting important variables
- Including irrelevant variables
- Using the wrong functional form
- Measurement error in an independent variable (and in the dependent variable)
- Sample selection bias

Other Model Misspecification Problems: Heteroskedasticity

- Heteroskedasticity is when the variance of the error terms is not constant
- Example: income as a function of years someone has worked for a company
- If we have heteroskedasticity, our estimated coefficients will still be unbiased but they won't be as precise and our standard errors may be incorrect
- More advanced statistical software can help correct for heteroskedasticity

Other Model Misspecification Problems: Heteroskedasticity

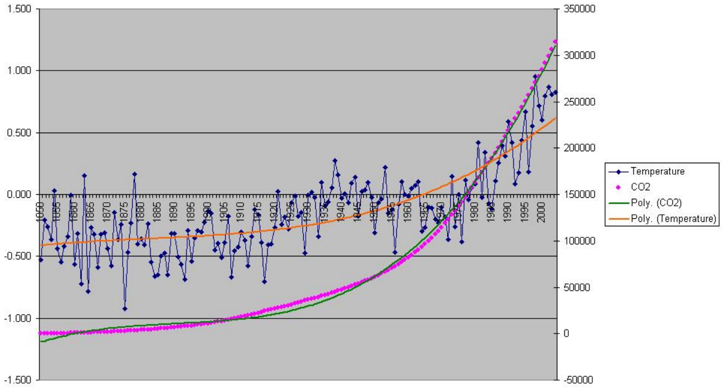


Other Model Misspecification Problems: Correlated errors

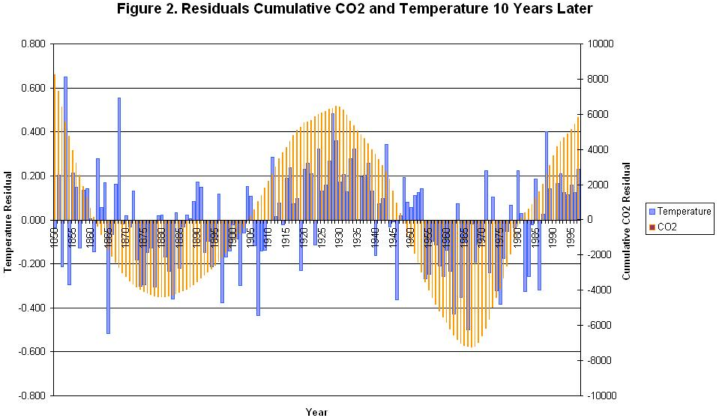
- Correlated errors: ε_i is correlated with ε_{i+1}
- This can often occur with time series data (if unemployment is higher than normal in one month, it will probably be higher than normal in the next month)
- It is also possible to have correlated errors in cross-sectional data (people from the same county may have similar unobservable characteristics, graduates of the same school may be more similar that graduates from different schools, etc.)
- Correlated errors complicate how we go about statistical inference

Other Model Misspecification Problems: Correlated errors

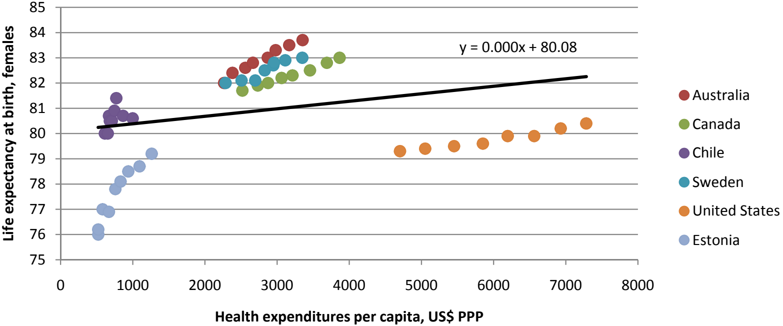
Figure 1. Cumulative CO2 Emissions and Temperature



Other Model Misspecification Problems: Correlated errors

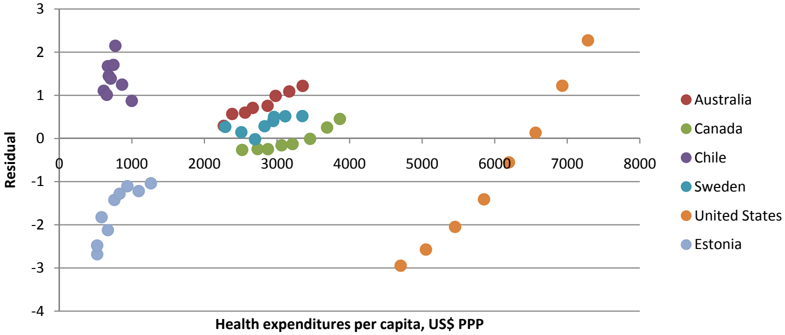


Other Model Misspecification Problems: Correlated errors



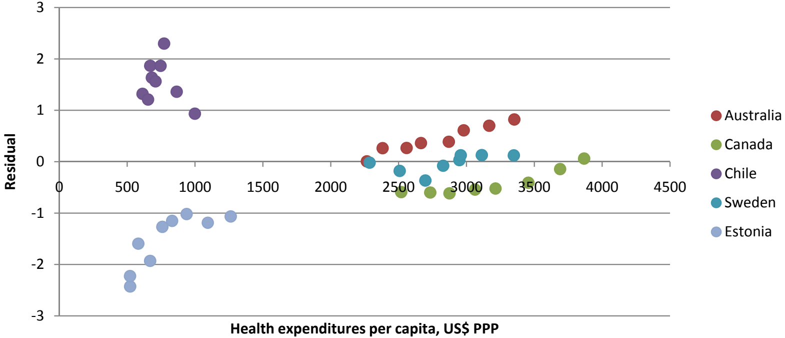
Data were obtained through the OECD StatExtracts system.

Other Model Misspecification Problems: Correlated errors



Data were obtained through the OECD StatExtracts system. Residuals are calculated as actual life expectancy minus predicted life expectancy using the results of a regression of female life expectancy on a quadratic in health spending per capita.

Other Model Misspecification Problems: Correlated errors



Data were obtained through the OECD StatExtracts system. Residuals are calculated as actual life expectancy minus predicted life expectancy using the results of a regression of female life expectancy on a quadratic in health spending per capita.

Other Model Misspecification Problems: Correlated errors

- Why are correlated errors a problem? Because we basically have less information than we think.
- Think of an extreme example, what if we just doubled our sample size by duplicating the dataset?
    - We'll get the same coefficient estimates but smaller standard errors (N is twice as big now)
    - But we've cheated somehow, we don't have any truly new information
    - The cheating shows up in the error terms, the information for each observation (including the error term) is perfectly correlated with the information of another observation in the dataset

Other Model Misspecification Problems: Correlated errors

- Now a less extreme example, what if we doubled our sample size by surveying two people in each household instead of just one?
  - We do get some new information but not as much as we might think
  - Unobservable characteristics will be correlated within households
  - Sampling two people at each of N households tells us less than sampling one person at each of $2N$ households
  - We need to take this into account when we calculate standard errors

Other Model Misspecification Problems: Correlated Errors

- So the main problem with correlated errors is that there is less information than a dataset with the same number of observations but uncorrelated errors
- With correlated errors we still get unbiased estimates of the slope coefficients but they will be less precise and the standard errors may be incorrect if we don't take this into account
- More advanced statistical software can help correct for correlated errors and give us correct standard errors

Other Model Misspecification Problems: Multicollinearity

- Multicollinearity occurs when we have a high degree of correlation between regressors (recall our parents' education example)
- Perfect collinearity:
  - Regressors are perfectly correlated
  - Estimation won't work, you need to drop one of the regressors
- Multicollinearity (not perfect):
  - Regressors are highly but not perfectly correlated
  - Estimation will work but standard errors will be really big
  - Estimates will be very sensitive to changes in the data

Moving From Association to Causality

- Everything we've developed so far still only addresses associations between variables, not causal links
- Even if we control for as many variables as possible, our estimated coefficients still do not tell us about causality
- There are a variety of techniques economists use to try to tease out causal relationships
- We'll take a brief look at a few approaches

Randomly Assigning Treatments

- One of the best ways for a social scientist to get at causality is to mimic other scientists
- In a lab setting, you might hold all relevant variables fixed and then change the variable of interest
- If you see a change in your dependent variable you can be pretty certain the change in the independent variable caused it
- It's tough to do this out in the real world
- One approach that is similar in spirit: randomize treatments

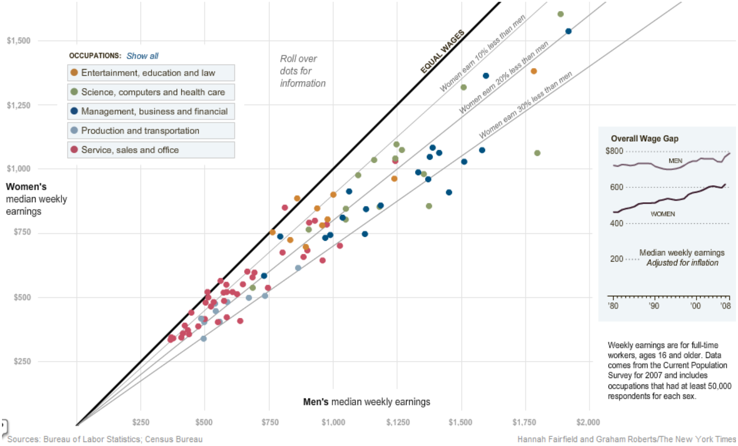
Social Assistance Programs: The New Hope Experiment

	Full Sample			One-Barrier Group		
Outcome	Program Group	Control Group	Difference	Program Group	Control Group	Difference
Percent of quarters employed (%)						
Years 1 to 3	72.7	67.2	5.5 ***	74.1	65.1	9.0 ***
Year 5	67.0	66.6	0.4	69.3	62.8	6.5 *
Year 8	56.3	54.2	2.1	60.1	46.7	13.4 ***
Average annual earnings ($)						
Years 1 to 3	9,756	9,259	497	10,380	8,518	1,862 ***
Year 5	11,961	11,795	166	12,766	10,891	1,875 **
Year 8	11,319	11,031	288	12,455	9,442	3,012 ***
Average records-based total income[a] ($)						
Years 1 to 3	14,971	13,921	1,051 ***	15,255	12,986	2,269 ***
Year 5	14,584	14,371	214	15,105	13,321	1,784 **
Year 8	13,595	13,285	311	14,458	11,472	2,986 ***
Total records-based income below the poverty standard[a] (%)						
Years 1 to 3	60.9	71.6	-10.7 ***	57.2	78.8	-21.6 ***
Year 5	59.3	64.6	-5.3 *	55.2	68.0	-12.7 ***
Year 8	63.1	67.1	-4.0	56.9	72.3	-15.3 ***
Sample size			1,357			580

Holding Everything Else Constant: Audit Studies

- Certain treatments can't be randomly assigned in this way
- Think about gender, we can't randomly switch the gender of study participants
- This is a problem because all sorts of characteristics and life experiences are correlated with gender
- When we try to study gender discrimination, it is tough determine whether differences in outcomes are due to discrimination or due to these other factors correlated with gender
- What if you could create people that looked identical except for their gender?

Gender Discrimination: Audit Studies



Gender Discrimination: Audit Studies

	# Audits (1)	No offers/ interviews (2)	Offer/inter. to both (3)	Offer/inter. to male only (4)	Offer/inter. to female only (5)	Male success (6)	Female success (7)	Test statistics h_0: no discrimination (p-value) Paired difference (8)	Symmetry (9)
A. High-price									
Offers	23	.48	.04	.43	.04	.48	.09	.01	.01
Interviews	23	.26	.13	.48	.13	.61	.26	.04	.02
B. Medium-price									
Offers	21	.43	.19	.29	.10	.48	.29	.18	.11
Interviews	21	.29	.33	.29	.10	.62	.43	.18	.11
C. Low-price									
Offers	21	.62	.10	.00	.29	.10	.38	.01	.02
Interviews	21	.52	.10	.10	.29	.19	.38	.18	.11

Neumark, Bank and Van Nort. "Sex Discrimination in Restaurant Hiring: An Audit Study" Quarterly Journal of Economics, 111 (3) 1996.

Natural Experiments

- Sometimes it's impossible or unethical to randomly assign treatments to people
- However, even if an economist can't randomly assign treatments, nature may be able to
- Consider trying to figure out the effect of having a larger family on the decision to work
- People choose their family size making family size correlated with preferences and characteristics that may also influence work decisions
- Economists look for a source of variation in family size that isn't due to these unobserved preferences and characteristics

Family Size and Work: Child Gender as an Instrument

Percentage of women having a third child by gender of first two children

Sex of the first two children	Fraction of sample	Fraction that had another child
one boy, one girl	0.494	0.372
two girls	0.242	0.441
two boys	0.264	0.423

From Angrist and Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" American Economic Review, 88(3), 1998.

Natural Experiments

- Often times these sources of random variation can come from the ways laws, regulations and programs work
- Consider class size and student performance
- People would like to know if larger classes lead to better or worse student performance
- The problem is, there are lots of things correlated with both class size and student performance that will bias our results (school district resources, funding, overall school district size, physical space constraints, etc.)
- So how can we distinguish the effect of class size from all of these other factors?

Class Size and Student Performance: Exploiting Maximum Class Size Rules