PAPERS DOC: a few ideas for paper topics

1. Contingent work:

(use the contingent worker supplement of the CPS.) Some workers are "leased" to a company, or only work on a "contract" basis in which their employment is understood to terminate when a project is over (usually, the company that they are working at does not pay them benefits--sometimes the company that leases them out does). These contingent workers include both some well paid professionals, but it also includes a lot of lower paid workers (janitors, food services, labor economics professors, etc.). You can see how contingent work (which can be defined in various ways) affects the receipt of UI benefits, Workers' Compensation benefits, the probability of being married, the likelihood of working in the service sector (these first four are probit or logit regression models), the level of wages, etc., by running one of these regressions:

UI benefits(Y=1,N=0)= β_0 + β_1 Cont Work(Y=1,N=0) + β_2 Other Stuff

WC benefits(Y=1,N=0)= β_0 + β_1 Cont Work(Y=1,N=0) + β_2 Other Stuff

Married (Y=1,N=0)= β_0 + β_1 Cont Work(Y=1,N=0) + β_2 Other Stuff

Service Industry Worker(Y=1,N=0)= β_0 + β_1 Cont Work(Y=1,N=0) + β_2 Other Stuff

Ave. Weekly Wage(continuous)= β_0 + β_1 Cont Work(Y=1,N=0) + β_2 Other Stuff

The estimated coefficient β_1 in each case would indicate how contingent work affected the dependent variable. For example, if β_1 were -24.56 in the Wage equation, it would indicate that contingent workers make $24.56 less per week than non-contingent workers, holding "other stuff" constant.

2. Discrimination:

Suppose you want to examine the empirical significance of the male/female wage gap (Lev. 27:3-4). We will employ a technique described as the Oaxaca (or Blinder) Wage Decomposition. We first estimate separate regressions for males and females (this could work for other groups: blacks vs. whites, baby boomers vs. younger cohorts, etc., BUT REMEMBER--YOU ONLY DO THIS FOR THOSE WHO ACTUALLY GET WAGES, i.e., no wage=0 people):

$\text{Wage}^M = \hat{\beta}_0{}^M + \hat{\beta}_1{}^M \text{ Education}^M + \hat{\beta}_2{}^M \text{ Professional}^M$ for males, and

$\text{Wage}^F = \hat{\beta}_0{}^F + \hat{\beta}_1{}^F \text{ Education}^F + \hat{\beta}_2{}^F \text{ Professional}^F$ for females.

Now for the decomposition, we will measure the variables at their sample mean values, and calculate the difference in Wage^M-Wage^F. Some of the difference in wages may be due to the level of education on average, or the industry or occupation that the worker is in (you should have a number of such variables, which we represent here with only a single dummy variable, Professional, to indicate whether the individual worker is employed in a professional position or not). You can show that, substituting in the estimated right hand side wage equation that we get:

$$\text{Wage}^M\text{-Wage}^F = \hat{\beta}_0^{\,M} + \hat{\beta}_1^{\,M}\ \text{Education}^M + \hat{\beta}_2^{\,M}\ \text{Professional}^M - (\hat{\beta}_0^{\,F} + \hat{\beta}_1^{\,F}\ \text{Education}^F + \hat{\beta}_2^{\,F}\ \text{Professional}^F)$$

$$= (\hat{\beta}_0^{\,M} - \hat{\beta}_0^{\,F}) + (\hat{\beta}_1^{\,M} - \hat{\beta}_1^{\,F})\ \text{Education}^M + (\hat{\beta}_2^{\,M} - \hat{\beta}_2^{\,F})\ \text{Professional}^M + \hat{\beta}_1^{\,F}\ (\text{Education}^M - \text{Education}^F) + \hat{\beta}_2^{\,F}\ (\text{Professional}^M - \text{Professional}^F)$$

The last two, right-most terms, capture the differences in wages due to educational and occupational differences, differences not attributable to "wage discrimination" but more likely represents differences in "productivity." (This is not an innocuous statement in the presence of occupational segregation.)

The three, left most terms:

$$(\hat{\beta}_0^{\,M} - \hat{\beta}_0^{\,F}) + (\hat{\beta}_1^{\,M} - \hat{\beta}_1^{\,F})\ \text{Education}^M + (\hat{\beta}_2^{\,M} - \hat{\beta}_2^{\,F})\ \text{Professional}^M,$$

represent differences in wages due to differences in the returns to schooling ($\hat{\beta}_1^{\,M} - \hat{\beta}_1^{\,F}$) differences in pay to Professionals ($\hat{\beta}_2^{\,M} - \hat{\beta}_2^{\,F}$), and differences in "initial wages" (the difference in intercepts represent differences in wages if there were no schooling and the workers were not professionals, say a defensive tackle playing for the U). These unexplained differences in wages is usually interpreted as discriminatory differences.

As an example, let:

$\text{Wage}^M = 600$ $\hat{\beta}_0^{\,M} = 125$

$\hat{\beta}_1^{\,M} = 25$ $\text{Education}^M = 15$ (on average, males have 15 years of schooling)

$\hat{\beta}_2^{\,M} = 1000$ $\text{Professional}^M = .10$ (10 percent of male workers are professionals)

$\text{Wage}^F = 400$ $\hat{\beta}_0^{\,F} = 115$

$\hat{\beta}_1^{\,F} = 20$ $\text{Education}^F = 13$

$\hat{\beta}_2^{\,F} = 500$ $\text{Professional}^F = .05$

Substituting these terms into the decomposition above (called either an Oaxaca or Blinder decomposition), we get:

\$200 = \$135 (due to "discrimination") + \$65 (due to "productivity"),

so that we estimate that two thirds the difference in wages is due to discrimination.

3. Labor supply of females

Labor supply of males is not as interesting, because there hasn't been as much variation as for females. So the dependent variable is usual hours worked per week (or weeks per year, or both: total hours per year):

$$\text{hrs} == \beta_0 + \beta_1\ \text{wages} + \beta_2\ \text{education} + \beta_3\ \#\ \text{children under age 6} + \beta_4\ \text{family income} + \beta_5\ \text{stuff} + (\beta_6\ \text{husband's wage-- if interested in family labor supply})$$

There is an interesting twist here, since not all women work there is a sample selection problem in which wages are only observed for those working (and those not working presumably are not working because they have lower market or higher reservation wages than those working).

One way to control for this sample selection is to use a two-stage procedure in which you first estimate a model of participation (and get a predicted probably of working for each individual), then in the second stage you include an additional variable (or variables) to control for the effect of sample selection. Here is an easy (but not quite right way to do it):

```
*coach's made up example estimate participation, then wage square
sample 1 20
read lfp hrs hrwage educ child6 fam_inc
0 0 0 13 4 43000
0 0 0 16 1 25000
0 0 0 12 0  49000
0 0 0 15 2 35000
0 0 0 11 2 53000
0 0 0 12 0 77000
1 40 20.50 12 0 38000
1 32 15.50 13 1 50000
1 6 7.50 10 0 20000
1 12 11.50 12 1 34400
1 28 12.50 13 0 90000
1 14 9.50 11 0 15000
1 22 8.50 14 0 37000
1 46 23.50 18 1 43000
1 40 12.75 16 3 67000
1 40 18.50 16 0 24000
1 29 13.80 12 0 53000
1 32 12.50 17 0 30000
1 22 9.50 16 1 85000
1 18 14.50 11 0 23000
if (hrs.eq.0) hrs=-99999
set skipmiss
* restrict sample to workers and predict wages for non-workers
sample 7 20
ols hrwage educ child6 fam_inc
fc /predict=hrwage beg=1 end=6
* run 1st stage estimates to get predicted prob of participation
sample 1 20
probit lfp hrwage educ child6 fam_inc / predict=predtd_p
genr predtdsq=predtd_p**2
* now run the labor supply model with predicted p, p-sq in it:
sample 7 20
ols hrs hrwage educ child6 fam_inc predtd_p predtdsq
```

```
end
stop
```

An alternative, more industry-standard way is given in the Shazam manual, v. 8, p. 288-290 (for one thing, it corrects the standard errors in the second stage, whereas the simple procedure above does not).

3. Unemployment or Workers Compensation Insurance Analysis

You could analyze the determinates of either the frequency of claims, or the severity (as measured by the dollar payments received last year), as a function of the expected benefit payments. For the frequency:

probit (or logit) model with Y: 1=received some workers' comp income last year
0=no such income last year
(everyone who is a FT worker in your sample should be in this regression)

Or you can estimate the severity of the injury (though on a state basis, the sample size for this second regression may be too small for many states):

OLS regression with Y: WC benefit income received last year
(only those who answered yes to receiving income should be in this regression)

To get principal independent variable of interest, the expected replacement rate (which is the expected weekly benefits divided by the usually weekly wage) you have to calculate the expected benefits from the given maximum weekly benefit (MAX), the replacement rate (RR), and the given minimum weekly benefit (MIN) established in each state by statute (I will put these on the class web site, so you can use them for your paper if you chose to write on this topic: UIXXparm.doc and WCXXparm.doc where XX is the year for which the parameters are in effect). You can calculate the expected benefit by using the following shazam code in the order written (given the weekly wage, wklywg):

```
genr benefit=RR*wklywg
if (RR*wklywg.le.MIN) benefit=MIN
if (RR*wklywg.ge.MAX) benefit=MAX
```

The other indep. variables (besides the replacement rate) in these regressions might be age, occupation(should create 6-10 dummies), race, union contract coverage, educational attainment, and industry (should create 6-10 dummies).

You can do exactly the same thing for UI frequency or severity as well (as an alternative)

4. Fringe Benefit Job Lock:

Do fringe benefits (health insurance, pensions) encourage job lock (implicit contract sort of stuff). So "job lock" will be measured by the probability of a quit or tenure with current firm (need the tenure supplement). (If there is a variable on the person's health, that should also be included in the analysis. You might expect a positive health insurance*poor health status intereaction on the tenure and probability of not quitting.

So the regression may be

Tenure with current employer= β_0 + β_1 employer provided health insurance(Y=1,N=0) + β_2 employer provided pension + β_3 other stuff (including the health stuff if available)

Probability that may quit this year = β_0 + β_1 employer provided health insurance(Y=1,N=0) + β_2 employer provided pension + β_3 other stuff

5. More hours if Quasi-fixed costs:
Does the existence of fringe benefits cause employers to offer longer hours of work to current employees rather than hire more new workers? (quasi-fixed cost stuff) You can proxy the extent of quasi-fixed costs by using dummy variables for employer provided pension benefits and employer provided health insurance (you probably want to run at least one regression restricted only to full time workers). So here is the regression

Hrs = β_0 + β_1 employer provided health insurance(Y=1,N=0) + β_2 employer provided pension + β_3 other stuff, or

Probability of Overtime = β_0 + β_1 employer provided health insurance(Y=1,N=0) + β_2 employer provided pension + β_3 other stuff (using a logit or probit model)

6. Human Capital: Education, Experience and Wages

This is the most frequent of all regressions (often it is the basis for the specification of the discrimination model). Usually the dependent variable is the natural logarithm of wages (ln(wage), which is usually the dependent variable in the discrimination studies as well even though you can do it either way), as follows:

ln(wage) = β_0 + β_1 experience + β_2 education + β_3 other

where β_1=return to experience (implicit, on-the-job training) and β_2=returns to schooling. IF β_2 were estimated to be .09, then the returns to schooling would be roughly 9 percent (under the Mincer/Becker model); another year of schooling increases wages by about 9 percent. If you choose to do this model, you should check out the returns to college education relative to the returns to high school, relative to post college education if there are enough data points. Also you should include a dummy variable for self employment, and interact it with the educational attainment variable. The estimated coefficient on the interaction should have what sign if schooling is pure signaling; vs. schooling as human capital enhancement.

7. Compensating Wage Differentials

The trick is to see whether wages systematically increase as workers are exposed to more unpleasant working conditions. Exposure to unpleasantness is not usually measured in the CPS data surveys, so you will have to merge data from elsewhere in order to test this model. You can do that for work injuries (by state and industry, and fatal injuries only by industry) by going to the following BLS Web site:

http://www.osha.gov/oshstats/work.html

and then to your specific state. Then match your data from the CPS to injuries on the Web site on the basis of industry. Typical shazam code to do this might be:

```
if (indcode.eq.46) injury=23.7
```

In this case, the CPS injury code 46 corresponds to an industry that had 23.7 lost time cases per 100,000.

You can also try to use the fatal injury rates, and match them in the same way, but fatal injury rates are not broken down by state.

8. Internet and Computer Usuage

Develop some sort of index of computer sophisfication, and see how well it predicts people's wages (and does the affect vary by age). To do this you might specify a model something like:

$$\ln(\text{wage}) = \beta_0 + \beta_1 \text{ experience} + \beta_2 \text{ education} + \beta_3 \text{ computer usage index} + \beta_4 \text{ age} + \beta_5 \text{ (computer usage index*age)}$$

Or try some interactions with computer usage with education.

9. Marital Search and Length of Marriage (search, or household prod. chapters)

(use the fertility and marriage history supplement). Longer search should result in better marital matches (less chance for divorce), just as it does for unemployed workers (or so a theory would go). You might figure our several ways to test the marital search model, and other aspects of household production and the division of labor, with this supplement. Here is one idea: among all of those who married a second time, did those with the longest wait before the end of the first marriage and the beginning of the second marriage, having the lowest chance of the second marriage ending in divorce. That is, the model would be something like this (this obviously restricts the sample to those with a second marriage):

$$\text{probability (2}^{\text{nd}}\text{ marriage ended in divorce)} = \beta_0 + \beta_1 \text{ (time between 1}^{\text{st}}\text{ and 2}^{\text{nd}}\text{ marriage)} + \beta_2 \text{ education} + \beta_3 \text{ number of children + stuff}$$

This should provide a start on some possible topics; if you know someone with access to a human resource data base for a company that would let you use it for the class, you may be so with my prior approval (this would be a database with information on each employee, including gender, wage, occupation, participation in 401K, promotions, job evaluations, wage, geographical location, workers compensation claims, etc.). That would be a good source for several other paper topics.