University of California, Davis
Department of Economics
Labor Economics

Date: June 28, 2006
Time: 3 hours
Reading Time: 20 minutes

PRELIMINARY EXAMINATION FOR THE Ph.D. DEGREE

This exam has two parts. You must answer the question in part I. In part II, answer 3 of the 4 questions.

University of California, Davis
Department of Economics
Labor Economics

Date: June 28, 2006
Time: 3 hours
Reading Time: 20 minutes

Part I. You must answer this question.

The "value of a statistical life" (VSL) is defined as the amount by which the average person would have to be compensated to accept an increase in the risk of death. Ashenfelter and Greenstone (2004) use a unique quasi-experiment to estimate the VSL: In 1987, the federal government permitted states to raise the speed limit on rural interstate highways from 55 to 65 miles per hour (mph). States that adopted the higher speed limit saw fewer hours of total travel (*hours*) but more fatalities (*fatalities*) on rural interstates, holding constant vehicle miles traveled (*vmt*). Using the wage rate to place a monetary value on the time savings associated with the increased mortality risk, they are able to arrive at an estimate of the VSL.

The simplest version of their model is given by:

$$\ln(hours)_{st} = \theta \ln(fatalities)_{st} + \beta \ln(vmt)_{st} + \alpha_s + \eta_t + \varepsilon_{st}$$

$$\ln(fatalities)_{st} = \pi_F 1[65 \text{ } mph \text{ } in \text{ } effect]_{st} + \lambda_F \ln(vmt)_{st} + \alpha'_s + \eta'_t + v_{st}$$

where $\ln$ denotes the natural log, *s* denotes state and *t* denotes year. The terms α_s, η_t, α'_s, and η'_t are fixed effects, and the variable $1[65 \text{ } mph \text{ } in \text{ } effect]_{st}$ is an indicator variable equal to 1 if the 65 mph speed limit is in effect in state *s* in year *t*, 0 otherwise. The error terms ε_{st} and v_{st} are potentially serially correlated; they are also potentially correlated with each other at a given point in time.

a. Under what formal conditions would the least squares estimate of θ be biased by selection on observables? By selection on unobservables?

b. As suggested by the graph below, states which adopted the 65 mph speed limit had unexpectedly low highway fatality rates in 1986.

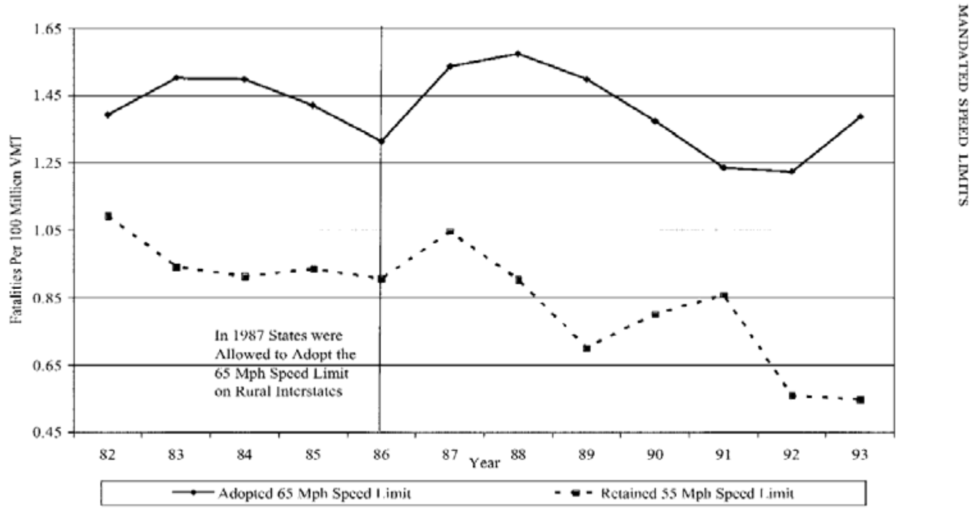


FIG. 3.—Trends in fatality rates on rural interstate roads, by adoption of 65-mph speed limit, 1982–93. The fatality rate is calculated as the weighted mean of the number of fatalities per 100 million VMT, where the weight is VMT.

S245

Would this tend to bias least squares estimates of π_F? If so, explain why and discuss ways in which this bias can be minimized.

c. Does π_F need to be identified to obtain a consistent two-stage least squares (TSLS) estimate of θ? Formally justify your answer.

d. An alternative way to implement this quasi-experiment is to use a "triple-difference" first stage model, where other "unaffected" road types (i.e., rural arterials, urban interstates) are used as an additional comparison group. Name at least one potential benefit and one potential drawback of this alternative model.

e. Below is a table showing first stage and reduced form coefficients on $1[65\ mph\ in\ effect]_{st}$ for alternative models. Based on the information in the table, would you say that TSLS estimates of θ are likely to be subject to a high degree of finite sample bias? Inconsistency? Justify your answer.

TABLE 5
TESTING THE ROBUSTNESS OF THE EFFECT OF THE 65-MPH SPEED LIMIT ON FATALITIES AND TRAVEL TIMES

Sample	(1)	(2)	(3)
	A. Dependent Variable: Ln(Fatalities)		
Rural interstates only	...	.360**	...
		(.091)	
Rural interstates and urban interstates	.312**	.417**	.414**
	(.097)	(.117)	(.130)
Rural interstates and rural arterials	.244**	.278**	.269**
	(.070)	(.099)	(.098)
All three road types	.280**	.349**	.337**
	(.073)	(.101)	(.096)
	B. Dependent Variable: Ln(Hours of Travel)		
Rural interstates only	...	−.041**	...
		(.007)	
Rural interstates and urban interstates	−.030**	−.032**	−.031**
	(.007)	(.010)	(.007)
Rural interstates and rural arterials	−.041**	−.040**	−.033**
	(.006)	(.010)	(.008)
All three road types	−.036**	−.036**	−.033**
	(.006)	(.009)	(.006)
ln(VMT) × road type	yes	yes	yes
State–road type indicators	yes	yes	yes
Year indicators	yes	no	no
Year–road type indicators	no	yes	yes
State-year indicators	no	no	yes

NOTE.—The two panels present results from regressions in which the dependent variables are the ln of fatalities and ln of hours driving, respectively. The entries are the estimated regression coefficients (heteroskedastic-consistent standard errors) for an indicator that is equal to one for observations from rural interstates when the 65-mph speed limit is in force. Road types are pooled for the analysis in different ways, as shown by the row labels. The bottom of the table lists the controls in each of the specifications. Number of observations is 326 for the rural interstates only sample, 653 for the rural interstates and urban interstates sample, 650 for the rural interstates and rural arterials sample, and 977 for the all-three sample.

* Significant at the 5 percent level.
** Significant at the 1 percent level.

University of California, Davis
Department of Economics
Labor Economics

Date: June 28, 2006
Time: 3 hours
Reading Time: 20 minutes

Part II. (Answer 3 of the following 4 questions.)

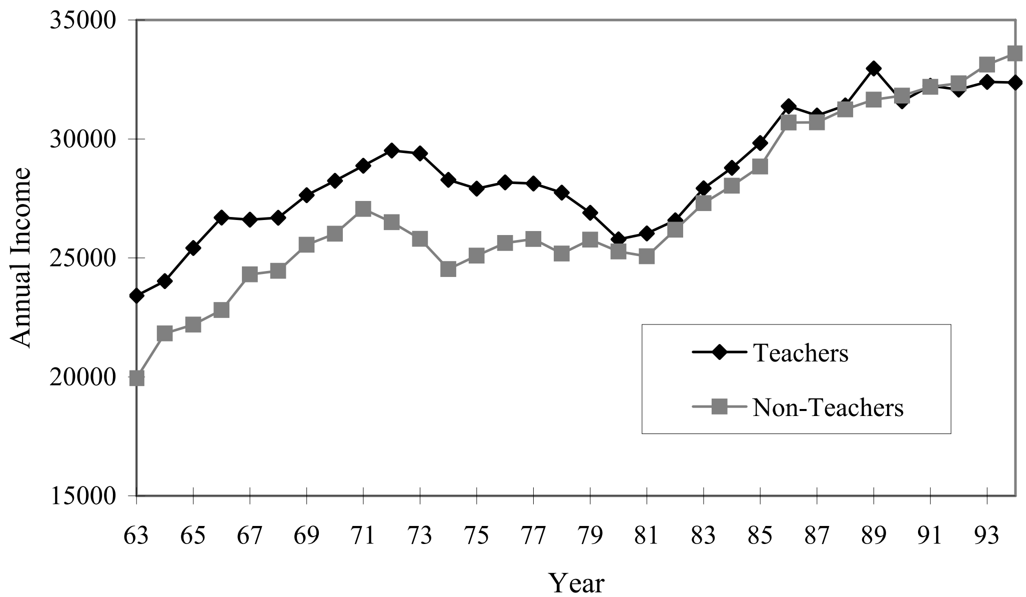
Q1. It is usually assumed that greater school inputs (such as higher paid teachers) lead to better schooling outcomes for children. However, empirical studies often do not find such a relationship. The starting point for the empirical literature on the effect of teacher wages is the following regression

$$Y_{is} = \alpha + \beta_1 TeacherWage_{is} + \beta_2 SchoolX_{is} + \beta_3 StudentX_{is} + \varepsilon_{is}$$

Where Y_{is} is an outcome for student *i* in school *s*, *SchoolX* is a vector of school characteristics and *StudentX* is a vector of student characteristics. The presumption is that β_1 is positive (i.e. children who go to higher quality schools will have better outcomes) but the empirical literature rarely finds that β_1 is positive and statistically significant. Most estimates, however, are based on cross-sectional data.

a. The justification for the positive relationship between teacher wages and student outcomes is based on a simple supply and demand graph, where the wage is the "teacher wage" and the quantity is "teacher quality." Draw this graph, and use it to explain why there should be a positive relationship between the teacher wage and student outcomes.
b. The most commonly acknowledged problem with estimating the above equation is that parents choose the schools that their children attend. Why would this affect the estimation of β_1? Would this problem likely lead to estimates that are too high, or too low? Explain your answer.
c. How might empirical researchers overcome this problem?
d. Another problem with estimating the above equation is that it makes no allowances for teachers' outside opportunities. What does the diagram below suggest about modeling the relationship between teacher wage levels and the supply of teachers? Could it be that failing to control for outside opportunities might help explain the frequent failure to find that β_1 is positive and statistically significant? Why?

Figure 2a
Mean Annual Wage and Salary Income (1994 dollars)



e. Another problem with estimating the equation is that there could be a compensating wage differential in the teacher labor market. In other words, some school districts might pay teachers higher wages in order to compensate them for other non-pecuniary conditions that make teaching in their school district less desirable than teaching in other school districts. How would this problem affect the estimates of β_1?

Q2. In their 2004 study, "Are Emily and Brendon More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," Bertrand and Mullainathan conducted an audit study to see whether employers discriminate against black job seekers. They selected 1,300 help-wanted ads from newspapers in Boston and Chicago and submitted multiple resumes from "phantom" job seekers. The researchers randomly assigned very "white" sounding names to half of the resumes and very African-American sounding names to the other half. Their main findings are reported in the two tables below.

a. What are their findings?
b. Do the results support the hypothesis that employers discriminate against black applicants, or does an alternative explanation better fit the results?
c. Given the evidence below, do the results support the idea of taste-based discrimination or statistical discrimination, or are they not distinguishable? Explain your answer and make sure that you describe the differences between these types of discrimination.

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (*p*-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Notes: The table reports, for the entire sample and different subsamples of sent resumes, the callback rates for applicants with a White-sounding name (column 1) an an African-American-sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these callback rates. In brackets in each cell is the number of resumes sent in that cell. Column 4 also reports the *p*-value for a test of proportion testing the null hypothesis that the callback rates are equal across racial groups.

TABLE 4—AVERAGE CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES AND RESUME QUALITY

	Panel A: Subjective Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*-value)
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)
	Panel B: Predicted Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Notes: Panel A reports the mean callback percents for applicant with a White name (row 1) and African-American name (row 2) depending on whether the resume was subjectively qualified as a lower quality or higher quality. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing the null hypothesis that the callback rates are equal across quality groups within each racial group. For Panel B, we use a third of the sample to estimate a probit regression of the callback dummy on the set of resume characteristics as displayed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). We then use the estimated coefficients on the set of resume characteristics to estimate a predicted callback for the remaining resumes (two-thirds of the sample). We call "high-quality" resumes the resumes that rank above the median predicted callback and "low-quality" resumes the resumes that rank below the median predicted callback. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing the null hypothesis that the callback percents are equal across quality groups within each racial group.

Q3. 1. In 1962, following Algeria's independence from France, 900,000 individuals of European origin returned to France from Algeria. Hunt (1992) examines the effects of this influx of repatriates from Algeria into France on unemployment and wages in France. Hunt relies on variation in the location of these repatriates across 90 departments (geographic areas, similar to provinces) throughout France for her empirical analysis. Results on the effects of the repatriates on unemployment of French non-repatriates (natives) are summarized in Table 3 (see next page)

a. In column 1 of Table 3, Hunt summarizes regressions of 1968 department level unemployment measures on the 1968 repatriates as a fraction of department labor force. Specifically, she estimates:

$$(\textit{nonrepatriate unemployed/nonrepatriate labor force})_{1968,\,i}$$
$$= \alpha_0 + \alpha_1 (\textit{repatriates} / \textit{laborforce})_{1968,i} + X_{1968,i}\beta + \varepsilon_{1968,i}$$

(where i indexes the departments) X includes controls for the age structure, education levels, and industry structure of the department.
What are the likely sources of bias (and their direction) on this OLS estimate? (View this regression as an attempt to estimate the impact of an exogenous influx of immigrants on the unemployment rate of natives.)

b. In column 2, Hunt uses the *1962* department level unemployment (measured just prior to the influx of repatriates) as the dependent variable, but the 1968 repatriates as the key independent variable. What does the coefficient on the 1968 repatriate variable suggest about the results in column 1 of this table? Explain.

c. Table 2 summarizes a regression relating the 1968 locations of the repatriates to a variety of department-level characteristics measured in 1962. Specifically these include the fraction of Algerian repatriates in the department prior to 1962, the average annual department temperature, and department average 1962 unemployment rates and annual salaries. What do these results suggest about the results discussed above? Do they suggest an alternative estimation strategy for identifying the effects of the repatriates on unemployment rates of non-repatriates? Explain such a strategy and discuss the extent to which it should eliminate the sources of bias mentioned in your answer to part a.

d. The demand for new housing construction increased dramatically from 1962 to 1965 in departments that received large numbers of Algerian repatriates. Use a simple supply and demand for labor graph to explain what this implies about the response to immigration estimated in this study.

(Hunt, 1992)

Table 3: Determinants of Unemployment Rates of Non-Repatriates

	Cross-Sectional Results	
	1968	*1962*
Independent Variable	*(1)*[a]	*(2)*[b]
Repatriates (% of 1968 Labor Force)	0.341** (0.089)	0.188** (0.053)
Age 15–24 (% of Labor Force)	−0.081* (0.043)	−0.030 (0.028)
Education (% with Bac.)	−0.251** (0.109)	−0.064 (0.073)
Services[c]	0.255** (0.066)	0.114** (0.042)
Commerce and Banking	−0.009 (0.046)	−0.006 (0.027)
Mining	0.061** (0.028)	0.019 (0.012)
Other Industry	0.006 (0.011)	−0.006 (0.005)
Construction	−0.072 (0.045)	−0.029 (0.025)
Public Sector	0.055 (0.043)	−0.006 (0.020)
Transport	0.126** (0.040)	0.092** (0.020)
Seven Regional Dummies	yes	yes
Adjusted R^2	0.83	0.79

Table 2. Determinants of the Location of Repatriates in the Labor Force, 1968. (Standard Errors in Parentheses)

Independent Variable	*Coefficient and Std. Error*
Repatriates from Algeria, 1954–62, as % of 1962 Population	2.44** (0.18)
Temperature	0.253** (0.048)
Unemployment Rate, 1962	0.325** (0.152)
Log Salary, 1962	−1.16** (0.54)

University of California, Davis — Date: June 28, 2006
Department of Economics — Time: 3 hours
Labor Economics — Reading Time: 20 minutes

Q4. Blundell, Duncan and Meghir (1998) (hereafter BDM) use a series of tax reforms in the UK to study labor supply behavior of married women.

a. One possible approach to estimating the labor supply effects of changes in net wages induced by tax reform is to compare changes over time in labor supply of tax-payers relative to non-taxpayers. BDM present such a comparison visually in their Figures 4 and 5, reproduced below. Why do BDM *NOT* rely on such comparisons in their study? What groups of individuals are used by BDM instead of the taxpayer/non-taxpayer grouping?

844 R. BLUNDELL, A. DUNCAN, AND C. MEGHIR

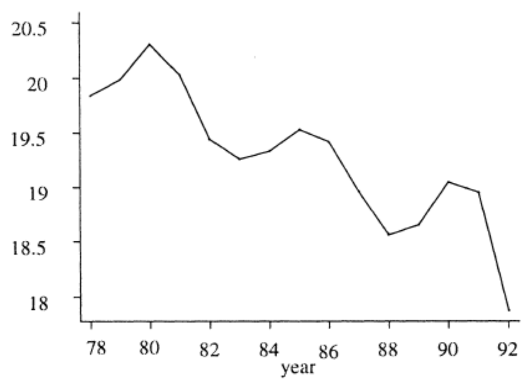


FIGURE 4.—Differences of female hours of work between taxpayers and non-taxpayers over time.

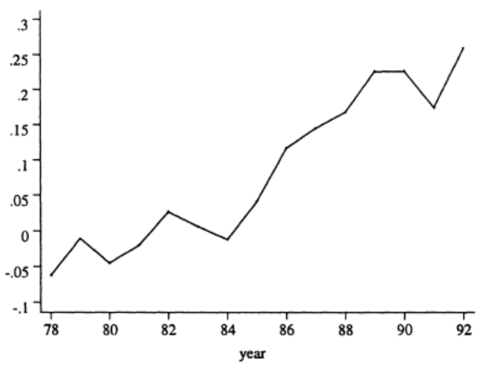


FIGURE 5.—Differences in female log wages between taxpayers and non-taxpayers over time.

b. BDM's main specifications are summarized by the following equation:

$$h_{it} = a_g + m_t + \theta' DK_{it} + \beta \ln w_{it} + \gamma \mu_{it}$$
$$+ \delta^w \hat{\nu}_{it}^w + \delta^\mu \hat{\nu}_{it}^\mu + \delta^P \hat{\nu}_{it}^P + \delta^T \hat{\nu}_{it}^T + e_{it},$$

Where h is hours of work for individual i in year t; a_g are group-specific fixed effects, m_t are year dummies, DK are demographic control variables, lnw is the log of after tax wages; μ is other (non-wage) income. The $\hat{\nu}$ terms are residuals from regressions of the log wage, other income, participation and tax kink indicators on group by time interaction terms.

What is the reason for including the wage residual terms?
Give an interpretation of the coefficients on the log wage, income, and wage residual terms from column (i) of Table VI. (below) What is the meaning of the change in the coefficient on the log wage in column (iv) of Table VI compared to the other columns of the table?

TABLE VI

ESTIMATES WITH NO DEMOGRAPHIC INTERACTIONS

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Constant	34.551	34.630	33.213	41.661	31.687	31.800
	3.386	*3.324*	*2.947*	*0.689*	*4.299*	*3.182*
DK02	−15.221	−15.211	−14.953	−13.079	−16.499	−16.492
	1.200	*1.200*	*1.137*	*0.509*	*1.397*	*1.074*
DK34	−16.033	−16.061	−16.112	−14.622	−16.945	−16.977
	1.214	*1.185*	*1.099*	*0.490*	*1.381*	*1.046*
DK510	−11.746	−11.774	−11.997	−11.025	−12.776	−12.805
	1.091	*1.067*	*0.971*	*0.325*	*1.233*	*0.945*
DK11O	−5.433	−5.443	−5.706	−5.118	−6.624	−6.632
	0.883	*0.875*	*0.794*	*0.347*	*1.039*	*0.810*
Log Wage	4.254	4.273	2.635	−2.446	2.851	2.894
	2.349	*2.341*	*2.054*	*0.346*	*3.062*	*2.265*
Other Income	−0.010	−0.010	0.004	−0.016	0.009	0.009
	0.015	*0.015*	*0.013*	*0.001*	*0.017*	*0.013*
Residuals						
Wage	−6.779	−6.795	−5.153		−7.371	−7.410
	2.405	*2.396*	*2.135*		*3.113*	*2.334*
Other Income	−0.006	−0.006	−0.020		−0.026	−0.026
	0.015	*0.015*	*0.013*		*0.017*	*0.013*
Tax Kink	0.337	0.332				
	0.076	*0.075*				
Participation	0.083				0.092	
	0.436				*0.356*	

c. The tax system in the UK includes a supplemental tax for national insurance contributions that generates a discontinuity in the tax schedule. Why is it important to control for this feature of the tax system when estimating labor supply elasticities? What is the likely effect on estimates of the wage coefficient (in an hours regression) if such kinks are not accounted for. Why?