POLICY ANALYSIS FIELD EXAMINATION QUESTIONS:
WINTER 2014

Question 1:

1. What assumptions are made in the Becker model about uncertainty and information about skills of workers? What does Becker assume about the mobility of capital in his two sector model of discrimination? In long run equilibrium in the Becker model of discrimination, what can be said about racial gaps in wages? Explain what is meant by the “marginal discriminator” and how the marginal discriminator differs from the average discriminator. Provide a graphic description of the short-run impacts of increases in the number of “most biased” discriminators in a Becker-type model, detailing your assumptions in constructing the graph.

2. What does (or would) the Phelps model say about the reliability of test scores in predicting true skill levels, ability, or productivity of whites versus blacks. Sketch the simple analytics of racial profiling using the Phelps model, making explicit the assumptions you make and detailing the results. In your model, is racial profiling efficient?

3. What is implicit discrimination and how does it differ from the two main economic models of discrimination (i.e. Becker’s taste for discrimination and statistical discrimination)? How does implicit discrimination differ from “spontaneous discrimination?” How would one distinguish empirically implicit discrimination from spontaneous discrimination?

4. List four major weaknesses or concerns about conventional regression models of discrimination (i.e. dummy variable method and Blinder-Oaxaca decomposition). What is Jeremiah Cotton’s main criticism with Blinder-Oaxaca decomposition and what does he recommend using to correct this issue?

5. You are given the following logistic regression model relating the probability of prequalification in construction as function of firm characteristics:

\[ \Pr(Y = 1) = \frac{1}{1 + \exp(- (\alpha + \beta_1 \text{DBE} + \beta_2 \text{Instate} + \beta_3 \text{Building} + \beta_4 \text{Heavy}) + \varepsilon)} \]

Where \( Y \) indicates whether a firm was pre-qualified, \( \text{DBE} \) indicates DBE status, \( \text{Instate} \) indicates whether the firm is headquartered inside the state or outside the state, \( \text{Building} \) represents whether the firm is in the sub-industry of building construction, and \( \text{Heavy} \) indicates whether the firm is in the sub-industry of heavy construction.

Table 1 presents the results of estimating the coefficients on the sample of firms that applied for prequalification in the previous year.
The organization of women and minority owned businesses claims that their members are less likely to be prequalified for construction work than are non-minority, non-women owned business enterprises.

Indicate wherein you agree or disagree based on the findings in Table 1. The organization claims that the evidence provided shows that the agency discriminates against DBEs. The agency responsible for prequalification hires you to defend it against allegations of discrimination. Provide a detailed methodology for responding to the allegations.

### Table 1: Logistic Regression Analysis of Pre-Qualification in Construction

<table>
<thead>
<tr>
<th></th>
<th>Coefficient Estimate</th>
<th>Chi-square</th>
<th>p-value</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.6736*</td>
<td>6.669</td>
<td>0.0098</td>
<td>1.0000</td>
</tr>
<tr>
<td>DBE Status</td>
<td>0.6493*</td>
<td>6.280</td>
<td>0.0122</td>
<td>0.1077</td>
</tr>
<tr>
<td>Instate</td>
<td>0.9889*</td>
<td>16.546</td>
<td>0.0001</td>
<td>0.8098</td>
</tr>
<tr>
<td>Building Construction</td>
<td>-0.3717</td>
<td>2.843</td>
<td>0.0918</td>
<td>0.3706</td>
</tr>
<tr>
<td>Heavy Construction</td>
<td>-1.5555*</td>
<td>46.119</td>
<td>0.0001</td>
<td>0.4448</td>
</tr>
<tr>
<td>Mean of Dependent</td>
<td></td>
<td></td>
<td></td>
<td>0.3678</td>
</tr>
<tr>
<td>Max-rescaled R-Square</td>
<td>0.1661</td>
<td></td>
<td></td>
<td>0.0001</td>
</tr>
<tr>
<td>Number of Observations</td>
<td></td>
<td></td>
<td></td>
<td>715</td>
</tr>
</tbody>
</table>

* Coefficient estimates are significant at 95% significance level.

Source: DOT Pre-Qualification of Construction File, 1999
Question 2:

1. Discuss the economic role of government in the provision of education, making sure to include efficiency and equity considerations in your answer.

2. Consider a simple linear regression model intended to estimate the effect of private vs. public education on an outcome such as test scores. For individual i, the model is:

   \[ Y_i = a + bX_i + cD_i + e_i \]

   where \( X \) is a vector of student or family specific characteristics, \( D \) is an indicator variable equal to one if the student attends private school, and \( e \) is an unobserved random error.

   a. Under what assumptions might the estimate of \( c \) in the equation above reflect the causal impact of private school attendance on test scores?

   b. Now describe the use of Instrumental Variable estimation to estimate the effect of private school attendance. What characteristics would a proper instrument satisfy?

   c. Instead of IV estimation, describe the use of Propensity Score Matching in estimating the effect of private school attendance. What specific problem with OLS does PSM try to address?

   d. Provide a general discussion of the advantages of disadvantages of IV versus PSM estimation for this estimation problem.
Question 3:

From *The Guardian*, January 16, 2014:
“A groundbreaking tax on sugar-sweetened beverages [and high-calorie processed foods] recently passed in Mexico could provide the evidence needed to justify similar laws across low- and middle-income countries and cities in the U.S., experts believe.

Campaigners and public health experts are watching closely to see what impact Mexico's tax has on consumption. Mexico, where 32.8% of the population is obese, is now the country with the biggest weight problem in the world, according to the UN's Food and Agricultural Organisation, overtaking the United States. The impact on health has been serious – 14% of the population has diabetes. Rates of high blood pressure, which can lead to stroke and heart attacks, are also high.”

Mexico’s new taxes include:

- an 8% tax on processed food that contains more than 275 calories per 100g and
- a one-peso (about $0.08) per liter tax on sugar-sweetened beverages.

In your responses to the following questions about the new Mexican taxes, include any models or literature you find relevant. Your answers should demonstrate an advanced graduate level of economics knowledge.

a) Can economic principles justify government intervention to reduce obesity? Under what circumstances?

b) Imagine that prior to enactment of the new taxes you are asked to evaluate their likely impact on equity. What information would you need to determine the likely incidence of the taxes? How would you predict the effect of the taxes on the distribution of the tax burden and on the distribution of after-tax income? Explain.

c) What tax compliance issues (if any) are likely to be raised by the new Mexican taxes?

d) Imagine that five years after enactment you are asked to measure the rate of non-compliance with the food and beverage taxes. What information would you need and what approach would you take to determine the rate of non-compliance?

e) Five years after the new taxes are enacted, how would you evaluate the policy’s effectiveness? Propose a study to determine whether the taxes have had the intended effect of reducing obesity. What data would you need? What approach would you use? Explain.

f) About 25 percent of adults in California are obese, and 10 percent of the state’s adult population has diabetes. If you were a California legislator, would you be in favor of enacting state taxes like Mexico’s high-calorie food and beverage taxes? Why or why not?
Question 4:

Interpret and critique the results of a study published in *Statistics in Medicine* (“Using the Peters-Belson Method to Measure Health Care Disparities from Complex Survey Data”) using the following table that included the following covariates: age, age², income, education (less than high school graduate, high school graduate, at least some college), region of residence (Northeast, West, Midwest, and South), metropolitan statistical area (in MSA or not), health insurance coverage (yes or no), and usual source of coverage (yes or no).

The Peters-Belson method assumes that the probability of a positive outcome for a person i in group R₁ should follow the same logistic model as a person from the reference group R₀, so each person i in group R₁ has a predicted outcome

\[ \hat{p}_{R₀,i} = \frac{\exp(x_i \hat{\beta}_{R₀})}{1 + \exp(x_i \hat{\beta}_{R₀})} \]

Thus, using the logistic model for group R₀, the estimated expected proportion of individuals with \( y=1 \) that are in group R₁ is

\[ \hat{p}_{R₀,R₁} = \frac{\sum_{j=1}^{n} w_j \delta_{R_i,j} \hat{p}_{R₀,j}}{\sum_{j=1}^{n} w_j \delta_{R_i,j}} \]

The estimated difference between the crude proportions of individuals in R₀ and R₁ with \( y = 1 \), \( p_{R₀} - p_{R₁} \), which is referred to as the overall disparity, can be partitioned as

\[ p_{R₀} - p_{R₁} = (p_{R₀} - \hat{p}_{R₀,R₁}) + (\hat{p}_{R₀,R₁} - p_{R₁}) \]

The difference \( p_{R₀} - \hat{p}_{R₀,R₁} \) estimates that part of the overall disparity that can be explained by the covariates and the difference \( \hat{p}_{R₀,R₁} - p_{R₁} \) estimates the ‘unexplained disparity’, which is that part of the overall disparity that cannot be explained by the covariates x in the logistic model. The proportion of the overall disparity that is not explained by the covariates may be due to discrimination or to omitted relevant covariates that are distributed differently between the groups. The proportion or percent of disparity that is explained by the covariates is estimated by \( D_{R₀}/D \) or \( 100 \times D_{R₀}/D \), which is referred to as ‘per cent explained’, where \( D = p_{R₀} - p_{R₁} \) and \( D_{R₀} = p_{R₀} - \hat{p}_{R₀,R₁} \).

Use this information along with the table below to answer the following questions.

A) Interpret and explain the differences in the percent explained across cancer screenings. What may contribute to the unexplained gap?

B) Does this method of computing the explained portion produce the same results as the Blinder-Oaxaca decomposition using STATA?

C) What can be said about the statistical significance of the explained portions displayed in column six?

D) What modifications, if any, would be required to apply this methodology to understanding gender differences in labor force participation?
Table I. Disparity in recent cancer screening* among women by race/ethnicity.

<table>
<thead>
<tr>
<th>Type of screening</th>
<th>Race</th>
<th>Sample size</th>
<th>Observed (per cent)</th>
<th>Predicted(^1) (per cent)</th>
<th>Per cent explained(^2)</th>
<th>Standard error of per cent explained (jackknife)</th>
<th>Standard error of per cent explained (Taylor linearization)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal (Age 50+)</td>
<td>White</td>
<td>4016</td>
<td>31.1</td>
<td>26.5</td>
<td>72.4</td>
<td>26.1</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>691</td>
<td>24.7</td>
<td>26.5</td>
<td>39.1</td>
<td>10.3</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>526</td>
<td>19.2</td>
<td>26.5</td>
<td>10.3</td>
<td>10.3</td>
<td>10.3</td>
</tr>
<tr>
<td>Digital rectal exam (Age 50+)</td>
<td>White</td>
<td>3993</td>
<td>43.4</td>
<td>38.5</td>
<td>93.4</td>
<td>47.6</td>
<td>47.2</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>683</td>
<td>38.2</td>
<td>35.0</td>
<td>75.7</td>
<td>21.5</td>
<td>21.5</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>518</td>
<td>32.3</td>
<td>35.0</td>
<td>21.5</td>
<td>21.5</td>
<td>21.5</td>
</tr>
<tr>
<td>Mammography (Age 40+)</td>
<td>White</td>
<td>5888</td>
<td>66.7</td>
<td>60.6</td>
<td>224.9</td>
<td>156.7</td>
<td>152.4</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>1049</td>
<td>63.9</td>
<td>60.6</td>
<td>154.3</td>
<td>47.7</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>928</td>
<td>60.1</td>
<td>56.6</td>
<td>47.7</td>
<td>47.5</td>
<td>47.5</td>
</tr>
<tr>
<td>Pap test (Age 25+)</td>
<td>Black</td>
<td>2354</td>
<td>80.1</td>
<td>77.6</td>
<td>79.2(^3)</td>
<td>26.6</td>
<td>26.5</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>11 077</td>
<td>77.0</td>
<td>77.6</td>
<td>1.8(^3)</td>
<td>8.8</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>2478</td>
<td>73.8</td>
<td>80.0</td>
<td>8.8</td>
<td>8.8</td>
<td>8.8</td>
</tr>
</tbody>
</table>

*For colorectal screening, ‘recent’ is if the respondent reported fecal occult blood test for screening during the past 2 years or endoscopy for screening during the last 3 years; for digital rectal exam and mammography, ‘recent’ is defined as during the past 2 years preceding the interview, and for Pap test, ‘recent’ defined as during the 3 years preceding the interview.

\(^1\)Predicted using the logistic regression model fit among the majority group (white) except for Pap test where the majority group was black.

\(^2\)((Observed (white) − predicted (black/Hispanic))/(observed (white) − observed (black/Hispanic)))\times100.

\(^3\)((Observed (black) − predicted (white/Hispanic))/(observed (black) − observed (white/Hispanic)))\times100.
Question 5:

Consider the question of whether to build a new public swimming pool in Dryville, which has a population of 70,230 people and a median household income of $31,500.

To understand the demand for swimming in a city pool, you have collected data from two dozen other towns. You decide to run a regression estimating the relationships between the number of VISITS per summer, the charge per visit (FEE), town population (POP) and median income (INCOME). The regression results are:

\[ \text{VISITS} = 140547 - 14638 \times \text{FEE} - 0.001127 \times \text{INCOME} + 0.6031 \times \text{POP} \]

a. Use the information in this problem to predict the total benefits of opening a public swimming pool in Dryville assuming that there will be free admission. Include a graph along with your numerical answer.

b. Predict total benefits if admission is set at $1.00.

c. At a price of admission price of $1, how much revenue would be raised? Calculate a number and show the relevant area in your graph. Also show and calculate the amount of consumer surplus that users of the pool in Dryville are likely to experience.

d. If the pool costs $500,000, would you recommend that Dryville should build a pool? Assume that both costs and benefits are only incurred for one year and there are no additional costs of disposal.
Question 6:

1. Describe the difference between a market good and a “Samuelsonian” pure public good or service (such as national defense). Show graphically how the determination of the optimal level of private goods and public goods differs. If all consumers have homogeneous tastes will all consumers demand the same quantity of the public good at any given price? Explain why or why not. Explain why the market result for providing pure public goods is likely to be sub-optimal. Is the market result likely to over-produce or under-produce the pure public good?

2. In a democracy one way of determining how much of the pure public good should be produced is by vote of elected representatives. Carefully explain whose preferences for public goods will be met using a simplified model in which there are 3 elected representatives and each representative has 5 constituents including herself, all of whom have different preferences for the public good. What assumptions are necessary for your model? Will the allocation of resources to the pure public good be optimal? Is the representative democracy model likely to over-produce or under-produce the pure public good?

3. Another way of determining how much of the pure public good should be produced is by popular vote or referenda. Assume a popularly elected executive defines the question to be decided as between service (and spending) levels A and B. Level A is the current level and is clearly less than optimal and level B is greater than level A. Assuming that the executive’s goal is to be re-elected and the executive has perfect information about the preferences of all constituents whose preferences will level B reflect?

4. If the executive has less than perfect information about voter preferences and voters believe the vote is a one-time only opportunity to change the level of services provided, is it rational for the electorate to approve a service level B which exceeds the level of services determined in part b above? Explain your result carefully. If voters believe there will be another referendum in the near term would the electorate vote to approve a service level that exceeds the level of services determined in part b? Explain why or why not?