219B – Final Exam – Spring 2009

Question #1 (CEO Overconfidence)

In this question we analyze the impact of overconfidence on corporate decision-making as in Malmendier and Tate (2005). Read carefully Table 4 from Malmendier and Tate (2005), including the Notes. The specification in Table 4 is a logit regression, taking the form

$$\Pr(Y_t = 1) = G\left(\alpha + \beta O_{it} + X_{it}B\right)$$

where G is a logistic distribution. The variable Long-holder is the proxy O of CEO overconfidence and it indicates that the CEO held a stock option until expiration rather than exercising it sooner. The dependent variable Y is an indicator variable for whether company i undertakes a merger in year t. Finally, notice that the coefficients are exponentiated, so that the coefficients can be read as odds ratios. Essentially, a coefficient of 1 indicates that the variable has no effect on the dependent variable, a coefficient bigger than 1 indicates a positive effect, etc. In parentheses are z-statistics (like t-statistics), not standard errors. (they should have been, but journal editors sometimes make extravagant requests)

a) Why is 'long-holder' a proxy of overconfidence? Describe the logic behind the choice of the authors. What form of overconfidence gives this prediction?

b) What is the intuition of why overconfident CEOs would undertake more mergers?

c) Assume now that overconfidence is of the 'overprecision' type: CEOs have on average the right prior about the value of their company, but they believe that their signal is too precise relative to the true uncertainty. Would you still predict a correlation between holding on to options and undertaking mergers?

d) Do the authors find a correlation between 'long-holder' and undertaking mergers? Use Column 2

e) What is an alternative interpretation of this correlation that does not rely on overconfidence? How would you distinguish it from overconfidence?

f) A couple of econometric questions: What does it mean that the standard errors are clustered by industry? Keep in mind that one observation is a firm-year combination. What kind of correlation of errors does this allow for? What kind of correlations does it rule out?

g) Right or wrong (explain your answer): 'Clustering the standard errors corrects the standard errors as well as the bias in the point estimates'.

h) How does the identification in Columns 2, 3, and 4 differ? In other words, how does the identifying variation change once one adds year fixed effects (Column 3) and Industry*Year fixed effects (Column 4)? In particular, what does it mean to add Industry*Year fixed effects? Why it is important to add such fixed effects? What happens to the result on overconfidence as one does so?

Table 4. Do Overconfident CEOs Complete More Mergers?

The dependent variable is binary where 1 signifies that the firm made at least one merger bid that was eventually successful in a particular firm year. Size is the log of assets at the beginning of the year. Q is the market value of assets over the book value of assets. Cash flow is earnings before extraordinary items plus depreciation and is normalized by capital at the beginning of the year. Stock ownership is the fraction of company stock owned by the CEO and his immediate family at the beginning of the year. Vested options are the CEO's holdings of options that are exercisable within 6 months of the beginning of the year, as a fraction of common shares outstanding. Vested options are multiplied by 10 so that the mean is roughly comparable to stock ownership. Corporate governance is a binary variable where 1 signifies that the board of directors has between four and twelve members.

Longholder is a binary variable where 1 signifies that the CEO at some point during his tenure held an option package until the last year before expiration. The fixed effects logit model is estimated consistently using a conditional logit specification. Standard errors in columns 1-3 are robust to heteroskedasticity and arbitrary within-firm serial correlation. Standard errors in column 4 are robust to heteroskedasticity and arbitrary within-industry correlation, where industries are measured using the 48 Fama and French industry groups (1997). Coefficients are presented as odds ratios.

	logit (1)	logit (2)	logit (3)	logit (4)	Random Effects logit (5)	Fixed Effects logit (6)
Size		0.9046	0.8733	0.8683	0.8600	0.6234
		(1.80)*	(1.95)*	(1.60)	(2.05)**	(2.60)***
Q_{t-1}		0.7719	0.7296	0.6651	0.7316	0.8291
		(2.85)***	(2.97)***	(2.37)**	(2.70)***	(1.11)
Cash Flow		1.9631	2.0534	2.1712	2.1816	2.6724
		(3.75)***	(3.93)***	(2.35)**	(3.68)***	(2.70)***
Stock Ownership		1.1212	1.2905	0.4126	1.3482	0.8208
		(0.13)	(0.30)	(0.67)	(0.28)	(0.11)
Vested Options		1.5912	1.5059	1.9596	0.9217	0.2802
		(2.56)**	(1.96)*	(1.46)	(0.19)	(2.36)**
Corporate Governance		0.6697	0.6556	0.6125	0.7192	1.0428
		(3.02)***	(3.08)***	(2.89)***	(2.17)**	(0.21)
Longholder	1.6831	1.5904	1.5557	1.5423	1.7006	2.5303
	(2.98)***	(2.72)***	(2.58)***	(1.90)*	(3.09)***	(2.67)***
Industry Fixed Effects	no	no	no	yes	no	no
Year Fixed Effects	no	no	yes	yes	yes	yes
Industry*Year Effects	no	no	no	yes	no	no
Observations	3690	3690	3690	2192	3690	2261
Number of Firms					327	184

Robust z statistics in parentheses. Constant included.

* significant at 10%; ** significant at 5%; *** significant at 1%

Question #2 (Employee Overconfidence)

Above, you discussed the evidence for overconfidence among CEOs. We now consider the possibility of overconfidence about own-company performance among rank-and-file employees.

a) A large share of rank-and-file employees in the US are paid stock options, in addition to a salary. These options pay if the stock of the company does well (that is, better than the issue price), and pay nothing otherwise. We analyze possible explanations for this phenomenon. A first possibility is that stock options are distributed to employees to incentivise them, similarly to how stock options incentivise the CEO and top managers. Is this a plausible explanation for why stock options are distributed to all employees in a company with thousands of workers? (United Airlines is a classical example) Discuss briefly.

b) A second possibility is that workers are overconfident about the performance of the company. Consider first a simple model with no overconfidence. At $t = 0$, United Airlines (UA) compensates its workers with $\$w$ wages and s units of stock options, with $w \geq 0$ and $s \geq 0$. Each stock option is worth \$1 if the stock price p_1 at $t = 1$ is higher than the price p_0 at $t = 0$, which occurs with probability π. The workers are risk-neutral and work for UA if their compensation package, which they value $w + \pi s$, is preferred to the alternative option $\bar{u} > 0$. The CEO maximizes profits, which are given by revenue R minus the wage bill w minus the value of the options granted (πs). In this simplified setting with no overconfidence, solve for the profit-maximizing combinations of w^* and s^* that the CEO of United Airlines sets. Compute the implied surplus for the workers and profits for the firm.

c) Assume now that workers are overconfident about the value of the firm and believe that the stock price will increase with probability $\tilde{\pi} > \pi$. The CEO instead is well-calibrated and believes that the probability of an increase in value is π, as in point b). Set-up the new maximization problem by the CEO and solve for the optimal w^* and s^*. Compute the implied surplus for the workers (both expected and true one) and profits for the firm.

d) Comment this statement: 'A common result in Behavioral IO when consumers have non-standard beliefs is that (i) consumers' utility in equilibrium is below their reservation utility; (ii) Firm profits are higher; (iii) The presence of non-standard beliefs affects the contract design' Relate to similar results in at least one paper in the Behavioral IO literature.

e) Consider now the case in which the workers are overconfident about the value of the firm ($\tilde{\pi} > \pi$), but the CEO is even more so: the CEO believes that the value of the option is $\bar{\pi} > \tilde{\pi} > \pi$, and hence maximizes $R - w - \bar{\pi} s$. Solve for the optimal w^* and s^*.

f) In light of your discussion in points b)-e), what form of overconfidence, if any, is consistent with the assignment of stock options to rank-and-file workers?

g) Building on points b) and c), assume now that the firms is hiring new workers, and workers are heterogeneous in overconfidence: a fraction q of workers is well-calibrated (they expect stocks to increase with probability π) and the remaining fraction $1-q$ is overconfident (the expect stocks to increase with probability $\tilde{\pi}$). Assume that there is a long line of workers out the door so there is excess supply of workers of each type. Assume that the CEO is, as in points b) and c), well calibrated. What contract will the firm now offer? Which types of workers will choose to apply to the company given the contract?

h) Use your results in g) to discuss this argument: 'Sorting can accentuate biases in the field'

i) In related work, Cowgill, Wolfers, and Zitzewitz (2007) document the betting market in Google. Google employees can bet on markets that forecast the completion time of Google projects (as well as other outcomes). In 2-outcome markets (i.e., 'The Google project XXX will be completed by August 30' vs. 'The Google project XXX will not be completed by August 30') the average price for a security that pays $1 if the favorable outcome occurs is 45 cents. The average frequency of the positive outcome occurring is 0.19. What does this imply about employee overconfidence?

Question #3 (Discounting and Crime)

We now consider a simple model of crime with (β, δ) preferences, and apply this model to empirical findings on deterrence effects. Ivan is deciding whether to commit a crime at time $t = 1$. Ivan lives in a 2-period model. If Ivan does not commit a crime, he obtains a payoff that we renormalize to zero at $t = 1$ and $t = 2$. If Ivan commits a crime, with probability p he is caught and goes to jail. In this case, the payoff is $u_j < 0$ (j for jail) in both periods, at $t = 1$ and $t = 2$. With probability $1 - p$, Ivan is not caught, consumes the stolen riches and he obtains payoff $u_r > 0$ (r for rich) at $t = 1$ and 0 at $t = 2$. (Ivan consumes all the riches in the first period).

a) Assume that Ivan is a (β, δ) discounter and compute the threshold level $\bar{p}$ of p for which Ivan is indifferent between stealing and not. Assume that the probability of being caught p is distributed with c.d.f. $F(p)$. Ivan observes the realization of p before deciding whether to commit a crime. Argue that the probability that Ivan commits a crime is $P(p \leq \bar{p}) = F(\bar{p})$. Substitute in the expression for $\bar{p}$ to obtain the probability of crime.

b) We now consider the deterrence effect, that is, how crime varies in response to harshness of sentencing. Compute how the probability of crime $F(p)$ varies with the severity of jail, that is, compute $\partial F(p) / \partial (u_j)$. (Do not forget that $u_j < 0$) Provide intuition on this comparative statics. Assuming a uniform distribution for p ($p \sim U[0, 1]$), compute $\partial^2 F(p) / \partial (u_j) \partial \beta$, that is, how the responsiveness of crime to u_j depends on β. How does the answer change for $\partial^2 F(p) / \partial (u_j) \partial \delta$?

c) Suppose that you can obtain perfect data on crime in a world like the one considered in a) and b). Can you identify separately β and δ from observed crime? If so, how? If not, why not?

d) We now modify the timing of payoffs in the setup above. Assume now that the benefits of stealing u_r are consumed half in period 1 and half in period 2. Nothing else in the setup changes. Repeat steps a) through b). What is the impact now of β and δ? Comment on the impact of time preferences on crime in this alternative version.

e) Now we add a period $t = 0$ to the crime game above. At time $t = 0$, Ivan votes whether to increase the jail cost in the future, that is, whether to make u_j more negative at $t = 1$ and $t = 2$. All the rest of the model is as in point a) and b). Ivan, who is a sophisticated present-biased agent, knows that his utility function as of period 0 is:

$$U = \beta\delta \left[\int_{\bar{p}} 0 dF(p) + \int^{\bar{p}} (p u_j (1 + \delta) + (1 - p) u_r) dF(p) \right]$$

where $\bar{p}$ is the threshold for committing crime that you derived above in point 1. Argue, formally if you can, that a time-consistent agent will never vote to make u_j more negative. Make an argument (intuitive is fine) that a present-biased (β, δ) agent may instead vote to lower u_j. (Do not attempt to find an example, but do provide intuition). Relate this to the paper by Gruber and Mullainathan on the effect of increased cigarette taxes on smoker happiness.

f) We now consider empirical tests of the prediction in Question b: longer prison sentences (more negative u_j) should lead to less crime. Consider a hypothetical empirical researcher that runs the following state-level panel regression:

$$crime_{j,t} = \alpha + \beta sentence_{j,t} + \delta_t + \theta_j + \varepsilon_{j,t}.$$

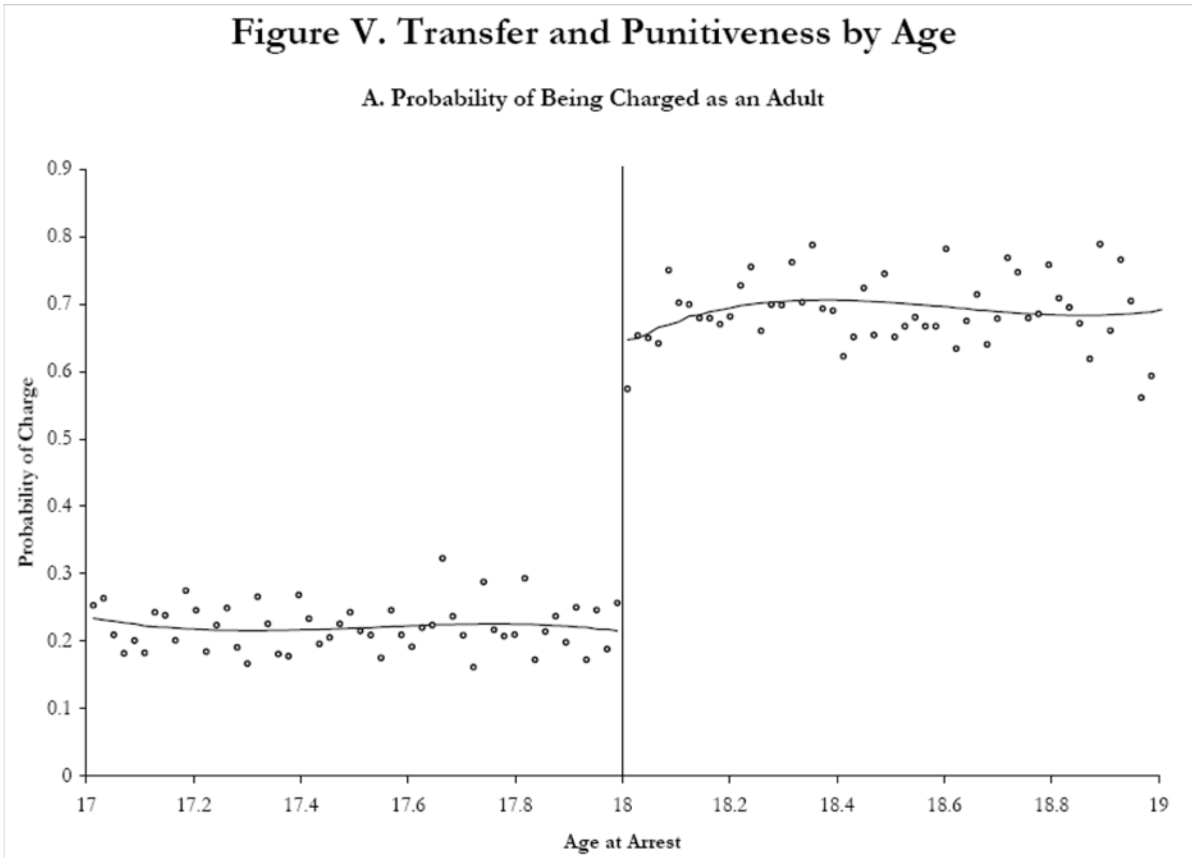
The dependent variable is the incidence of property crime (number of accidents/population) in state j in year t. The independent variable is the average sentence length (in weeks) for a property crime episode in state j in year t. Provide an interpretation for the coefficient β, given that year dummies δ_t and state dummies θ_j are included in the regression. What kind of variation identifies β?

g) The researcher finds the surprising result

$$\begin{aligned} crime_{j,t} &= \underset{(.0001)}{.0003} + \underset{(.003)}{.01} * sentence_{j,t} + \delta_t + \theta_j + \varepsilon_{j,t}. \end{aligned} \tag{1}$$

Tougher sentencing increases crime! Discuss at least one bias that can generate a spurious positive correlation. The researcher running (1) replies that the fixed effects for state and year should control for this. Counter-reply.

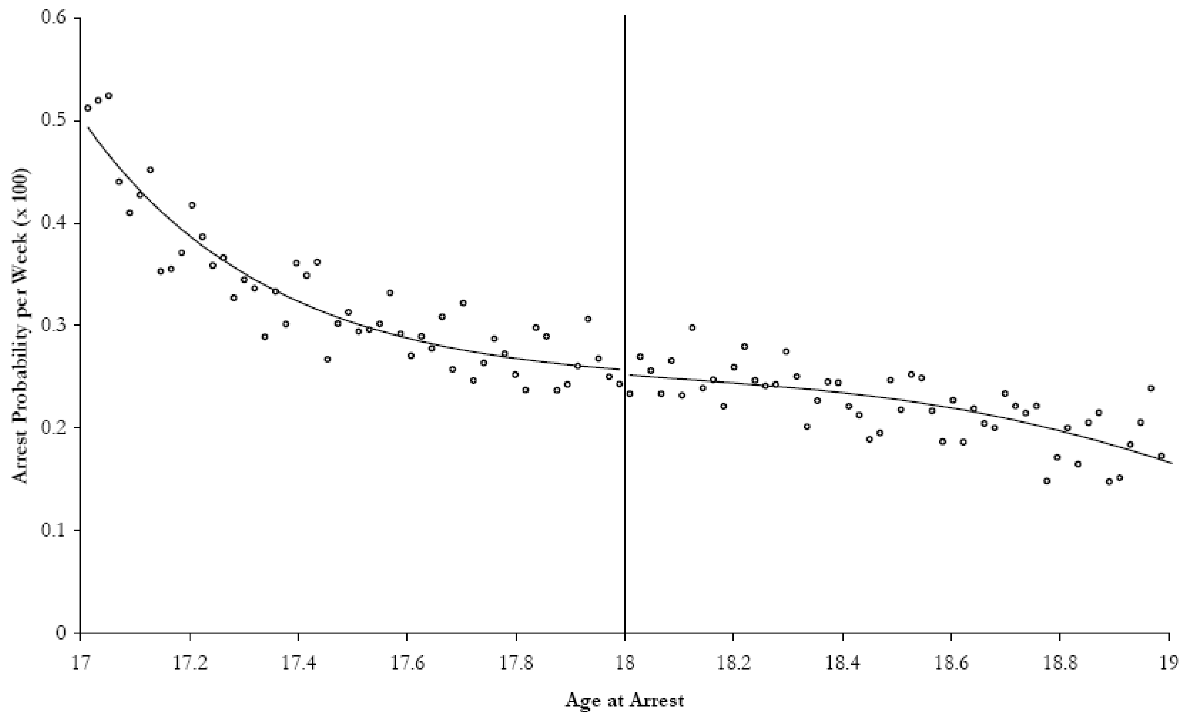
h) David Lee and Justin McCrary pursue a different venue to test the prediction in a). They compare the crime rate for individuals that are barely below 18 years of age to the crime rate of individuals that are barely above 18 years of age. In doing so, they exploit the discontinuity of sentence harshness upon turning adults. The average incarceration length for adults in their (Florida administrative) data is 3 to 7 times higher for adult arrests than for juvenile arrests. Figure V.A from their paper shows plots of the probability of adult incarceration as a function of the age at the time of the crime. Committing a crime right after age 18 (as opposed to right before 18) increases the probability of adult incarceration from about .2 to about .6. Given this result, the authors then propose the following test of deterrence: Is the likelihood of committing a crime lower for people just above 18 compared to people just below 18? Explain why this identification strategy does not suffer from the problems of the panel regressions in (f) and (g). Can you think of one potential issue for this identification strategy?



i) Figure IV.A plots arrest probability as a function of the exact age of the criminal. Surprisingly, there is no discontinuity around age 18 of arrest rates. This is the main finding of the paper. We now relate this finding to points (a) and (b). Consider the expression for $\bar{p}$ in (a), assume a return from stealing u_r of \$500 and per-period jail cost u_j of \$500 (capturing forgone earnings and leisure). Assume $p \sim U[0,1]$, $\delta = 1$ and $\beta = 1$ (standard model). How much does the probability of crime $F(\bar{p})$ increase if u_j doubles to \$1000 (as a function of β and δ)? How does this match the empirical findings?

Figure IV. Criminal Propensity Estimates by Age

A. First Arrest Since 17



j) What if you allow for present-biased preferences, with $\delta = 1$ and $\beta = .5$? Is present bias a reasonable explanation for the empirical findings (you may find the answer to question (d) also useful)?

Question #4 (Disposition Effect)

a) Odean (1998) provides evidence on the disposition effect, that is, the tendency of investors to sell losers rather than winners. The main finding is that the propensity to realize gains (PGR, 0.148) is significantly higher than the propensity to realize losses (PLR, 0.098). Using the numbers in the notes to the Tables, describe how the PGR and PLR are computed, you should be able to obtain the numbers $PGR = 0.148$ and $PLR = 0.098$.

Table I

PGR and PLR for the Entire Data Set

This table compares the aggregate Proportion of Gains Realized (PGR) to the aggregate Proportion of Losses Realized (PLR), where PGR is the number of realized gains divided by the number of realized gains plus the number of paper (unrealized) gains, and PLR is the number of realized losses divided by the number of realized losses plus the number of paper (unrealized) losses. Realized gains, paper gains, losses, and paper losses are aggregated over time (1987–1993) and across all accounts in the data set. PGR and PLR are reported for the entire year, for December only, and for January through November. For the entire year there are 13,883 realized gains, 79,658 paper gains, 11,930 realized losses, and 110,348 paper losses. For December there are 866 realized gains, 7,131 paper gains, 1,555 realized losses, and 10,604 paper losses. The t-statistics test the null hypotheses that the differences in proportions are equal to zero assuming that all realized gains, paper gains, realized losses, and paper losses result from independent decisions.

	Entire Year	December	Jan.–Nov.
PLR	0.098	0.128	0.094
PGR	0.148	0.108	0.152
Difference in proportions	−0.050	0.020	−0.058
t-statistic	−35	4.3	−38

b)Why is this pattern weaker reversed in December? Briefly explain the tax reasons to sell losers, and why this makes the disposition effect a puzzle.

c) Draw a picture of the prospect theory value function and discuss why prospect theory can in principle explain the disposition effect. Which feature (or features) of prospect theory helps to explain the disposition effect: (i) loss aversion; (ii) diminishing sensitivity (that is, concavity over gains and convexity over losses) ; (iii) non-linear probability weighting. Assume here and in what follows that the reference price is the initial purchase price. Provide intuition.

d) Elaborating on the previous point, does the simple piece-wise linear reference-dependent utility used in most of the literature

$$v(x) = \begin{cases} x - r & \text{if } x \geq r; \\ \lambda(x - r) & \text{if } x < r, \end{cases}$$

predict the disposition effect?

e) Building on Barberis and Xiong (forthcoming), explain why a more careful look at prospect theory may lead us to conclude that prospect theory does *not* explain the disposition effect. Provide as much intuition as you can on this difficult paper.