219B – Final Exam – Spring 2010

Question #1 (Confusion for voters). As we discussed in class, Kelly Shue and Erzo Luttmer estimate the role of confusion on choices of voters in the California recall elections of 2003. The ballot listed over 100 candidates, but only 3 obtained a significant share of the vote: Schwarzenegger, Bustamante, and McClintock. The order of candidates in the ballot is randomized and varies (randomly) across Assembly Districts. Hence, Kelly and Erzo decide to use this experimental variation to estimate confusion as the share of voters that minor party candidates obtain when they are placed near a major party candidate, relative to when they are not.

They estimate the specification

$$(VoteShare_i) * 100 = \beta_0 + \beta_1 * VoteShareAdjacent_j + Controls + \varepsilon$$

where i indicates the minor party candidate and $VoteShare_i$ his/her vote share, and $VoteShareAdjacent_j$ is zero if the minor party candidate is not adjacent to a major party candidate, and otherwise it equals the vote share of the major party candidate. They obtain the following results.

Table 2: Primary Results

Dependent Variable: *Voteshare* = (votes / total votes)×100	(1)	(2)	(3)
Adjacent	0.104** (0.018)		
Adjacent* × *Schwarzenegger		0.088** (0.025)	
Adjacent* × *Bustamante		0.143** (0.025)	
Adjacent* × *McClintock		0.107* (0.045)	
Adjacent Dummy			0.037** (0.006)
Observations	1,817,904	1,817,904	1,817,904
R-Squared	0.8676	0.8676	0.8676

a) Focus first on Column (1). Interpret the result as confusion explaining precisely what the regression coefficient captures. In light of this interpretation, how common does confusion appear to be? Compute a *confusion rate* as follows: the share of voters that meant to vote for a major party candidate that instead voted for another candidate. (Keep in mind that (i) the dependent variable is the vote share multiplied by 100, while the right-hand side variable is the actual vote share, and (ii) each major party candidate is surrounded by about 2.5 minor party candidates)

b) Consider now the results in Column (2) where the regression allows the coefficient β_1 to differ for the three main candidates. Provide an interpretation to the difference in coefficient between Schwarzenegger (Republican) and Bustamante (Democrat). How would you test directly your interpretation?

c) In Column (3) the authors estimate the specification

$$(VoteShare_i) * 100 = \beta_0 + \beta_1 * dAdjacent_j + Controls + \varepsilon$$

where $dAdjacent_j$ is an indicator variable for a candidate adjacent to one of the major 3 party candidates. Reconcile (approximately) the parameter estimate in Column (3) with the estimate in Column (1).

d) Provide at least one alternative interpretation of the result above that is not based on voter confusion.

e) Describe and comment on the results of Table 3, Columns (3), (4), and (6). Column (3) refers to the position of the minor party candidate relative to the major party candidate, and Column (6) exploits the fact that in a punch-card it is much easier to confuse the candidates and vote to the candidate to the left of the intended candidate, rather than to the right.

Table 3: Robustness Checks

Dependent Variable: *Voteshare* = (votes / total votes)×100	(1)	(2)	(3)	(4)	(5)	(6)
Adjacent	0.082**			0.104**	0.113**	
	(0.027)			(0.018)	(0.018)	
Adjacent Dummy	0.010					
	(0.007)					
Adjacent Dummy × CA Voteshare		0.112**				
		(0.019)				
North Adjacent			0.082**			0.082**
			(0.022)			(0.022)
South Adjacent			0.111**			0.111**
			(0.033)			(0.033)
East Adjacent			0.143**			
			(0.035)			
West Adjacent			0.038**			
			(0.011)			
Diagonally Adjacent				0.002		
				(0.003)		
Punchcard Adjacent					0.030+	
					(0.018)	
Horizontally Adjacent						0.031**
						(0.008)
Horizontally Adjacent × Confusing Side						0.123**
						(0.038)
Observations	1,817,904	1,817,904	1,817,904	1,817,904	1,817,904	1,817,904
R-Squared	0.8676	0.8676	0.8677	0.8676	0.8677	0.8677

f) What does this additional evidence suggest regarding the alternative interpretation you provided in point d)?

g) In all the above specifications, the standard errors are clustered at the county*district level. What kind of correlation does this clustering allow? Given an example of a possible reason for such intra-cluster correlation? What kind of correlation does this clustering forbid?

h) Why are the standard errors clustered at the county*district level? [Remember: The order of candidates differs within each county-district combination] Finally, would it be more restrictive to cluster the standard errors at the district level? Explain.

Question #2 (Confusion for investors).

a) Now consider the evidence on confusion from financial markets (Rashes, *JF*). The telephone company known as 'MCI' actually has ticker MCIC, while a much smaller investment company (Massmutual Corporate Investors) has MCI as ticker, hence the possible confusion. Comment on the pattern of correlations in daily trading volume (roughly speaking, the number of shares traded) in the next Table III in light of the confusion story. Interpret also the correlation between investment company (MCI ticker) and AT&T (T ticker).

Table III

Daily Volume Correlation Coefficient Matrices

This table presents the correlation of daily volumes between Massmutual Corporate Investors fund (MCI), MCI Communications (MCIC), AT&T (T) and the New York Stock Exchange Composite Index (NYSE). The pairwise Pearson product-moment correlations are shown with the standard error of these coefficients in parentheses.

	MCI	MCIC	T	NYSE
	Panel A: Sample Period 11/21/94–11/13/97			
MCI	1			
MCIC	0.5592	1		
	(0.0302)			
T	0.0291	0.1566	1	
	(0.0364)	(0.0360)		
NYSE	0.1162	0.2817	0.3397	1
	(0.0362)	(0.0350)	(0.0343)	

b) Next, the author present evidence on the correlation of daily returns between the MCI and the MCIC stock. The specification is

$$r_{MCI,t} = \alpha_0 + \alpha_1 r_{MCIC,t} + \beta X_t + \varepsilon_t. \tag{1}$$

Comment the results in the second row and third row of the next Table, including the magnitude of the α_1 coefficient. (Be careful: The specifications are in the rows, not in columns as usual. So the second row of coefficients captures the estimates for the specification in (1) with market controls, and the third row adds in addition the AT&T returns.

Constant	MCIC Return	(MCIC Return) * dummy (MCIC return <0)	T Return	S&P 500 Return	S&P Smallcap Return Residual	Lehman Long Bond Index Return	R^2
Panel A: Sample Period 11/22/94–11/13/97							
0.0956				0.0372	0.1011	0.0932	0.0286
(2.6223)				(0.9370)	(1.9233)	(2.3438)	0.0247
0.0954	0.0862			0.0128	0.1068	0.0905	0.0353
(2.6243)	(2.2779)			(0.3128)	(2.0356)	(2.2818)	0.0301
0.0957	0.0851		0.0171	0.0052	0.1077	0.0907	0.0355
(2.6306)	(2.2430)		(0.4190)	(0.1166)	(2.0501)	(2.2862)	0.0290
0.0721	0.1205	−0.0722		0.0149	0.1070	0.0913	0.0360
(1.5202)	(2.0557)	(−0.7664)		(0.3630)	(2.0375)	(2.3015)	0.0296

c) An economist summarizes the results as follows: 'Yes, there is some confusion as apparent from the volume of trades, but the evidence is a lot less strong for returns, given arbitrage'. Do you agree or disagree? Comment, documenting your thoughts.

d) Interpret the results in the last row which allows for a different effect depending on whether MCIC had a positive or negative return. What do the estimated coefficients imply is the effect of a 10 percent negative MCIC return on the MCI return?

e) Finally, you are not satisfied with just documenting qualitative patterns in two different data sets, you are looking for ways to compare the amount of confusion for voters and investors. Using the correlations in the volume data, can you compute a *confusion rate* to compare to the rate in the voter data? Define it as the share of investors that meant to trade the telephone company and instead traded the investment company. (Assume no error the other way) You would obtain it as the parameter β in the regression

$$V_{MCI,t} = \alpha + \beta V_{MCIC,t} + \varepsilon_t.$$

Unfortunately, Rashes does not report this regressions in the paper. However, can you still obtain an approximate estimate of the confusion rate $\hat{\beta}$? Use the correlation coefficient $\hat{\rho}$ in Table III (.5592) and the standard deviation of volume in the next Table (Hint: By the definition of ρ, $\rho_{MCI,MCIC} = Cov(V_{MCI,t}, V_{MCIC,t})/\sqrt{Var(V_{MCI,t})Var(V_{MCIC,t})}$. Use the expression for β from a univariate OLS regression.)

Summary Statistics

Daily return and volume information is shown for Massmutual Corporate Investors fund (MCI), MCI Communications (MCIC), and AT&T (T) for the sample period 11/21/94–11/13/97. The return for security j is expressed in percentages and defined as $\text{Log}[(P_{j,t+1} + D_{j,t+1})/P_{j,t}]$, where $P_{j,t}$ and $D_{j,t}$ are the price and dividend, respectively, for security j on day t.

	Mean (Return)	*SD* (Return)	Mean (Volume)	*SD* (Volume)	Mean (Price)
MCI	0.078	0.7136	4,155	4,497	36.14
MCIC	0.087	2.3645	4.154×10^6	4.713×10^6	28.07
T	0.055	1.6440	4.810×10^6	2.837×10^6	38.64

f) How does the confusion rate for investors compare to the one for voters?

Question #3 (Noise Traders).

In the previous question, we examined a specific case in which 'noise traders' that make a mistake in investment decisions can affect asset prices. Here we review, similarly to what we did in your problem set, the noise trader set-up (DeLong, Shleifer, Summers, Waldman, *JPE* 1990). There is a share μ of noise traders, $(1-\mu)$ of arbitrageurs. The arbitrageurs are risk averse and have a short horizon, that is, they have to sell the shares at the end of period to consumer. Formally, consider an OLG model where in period 1 the agents have initial endowment and trade, and in Period 2 they consume. There are two assets with identical dividend r: a safe asset with perfectly elastic supply, whose price we will set to 1 (numeraire), and an unsafe asset in inelastic supply (1 unit) and a price p that is determined by supply and demand. We denote the demand for unsafe asset: λ^a and λ^n. The investors have CARA utility function $U(w) = -e^{-2(\gamma w)}$ with w being the wealth in Period 2, which is what the investor consumes. Compared to the arbitrageurs, the noise traders believe that in period t the asset with have higher return ρ_t. Given that the wealth w is distributed $N\left(\overline{w}, \sigma_w^2\right)$, maximizing $EU\left(w\right)$ is equivalent to maximizing $\overline{w} - \gamma\sigma_w^2$, that is, the problem reduces to one of mean-variance optimization. [You do not need to show this]

a) Show that arbitrageurs maximize the problem

$$\max(w_t - \lambda_t^a p_t)(1+r) + \lambda_t^a(E_t[p_{t+1}] + r) - \gamma\left(\lambda_t^a\right)^2 Var_t(p_{t+1}).$$

Derive the first order condition and solve for λ_t^{a*}.

b) Show that noise traders maximize the problem

$$\max(w_t - \lambda_t^n p_t)(1+r) + \lambda_t^n(E_t[p_{t+1}] + \rho_t + r) - \gamma\left(\lambda_t^n\right)^2 Var_t(p_{t+1}).$$

Derive the first order condition and solve for λ_t^{n*}.

c) Discuss how the optimal demand of the risky asset will depend on the expected returns $(r + E_t[p_{t+1}] - (1+r)p_t)$, on risk aversion (γ), on the variance of returns $(Var_t(p_{t+1}))$, and on the overestimation ρ_t.

d) Under what conditions noise traders hold more of the risky asset than arbitrageurs do?

e) To solve for the price p_t, we impose the market-clearing condition $\lambda^n\mu + \lambda^a\left(1-\mu\right) = 1$. Use this condition to solve for p_t as a function of $E_t[p_{t+1}]$, $Var_t(p_{t+1})$, and the other parameters.

f) To solve for the equilibrium, assume that the average price is not time-varying (that is, $E_t\left[p_t\right] = E_t\left[p_{t+1}\right] = E[p]$), and take expectations on the right and left of the expression for p_t. Solve for $E\left[p\right]$, and substitute into the expression for p_t. Now, use this expression to compute $Var\left[p_t\right]$. Finally, substitute the expression for $Var\left[p_t\right]$ in the updated expression for p_t. In the end, you should obtain

$$p_t = 1 + \frac{\mu(\rho_t - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{2\gamma\mu^2\sigma_\rho^2}{r(1+r)^2}.$$

[If you get stuck, move on to the next point.]

g) Analyze how the price p responds to an increase in μ, in ρ_t, in ρ^*, in γ, and in σ_ρ^2. For each of these terms provide intuition.

i) In the dot-com bubble period (late 90s), there was a substantial inflow of individual investors in asset markets. Speculate in light of the above expression what the effect is likely to have been. State all assumptions you are making.

Question #4 (Present Bias and Procrastination of Retirement Savings).

In this Question we consider the impact of self-control problems on procrastination. Consider a present-biased individual that is considering when (and whether) to undertake an investment activity with immediate costs and delayed benefits. The main example will be calling the Human Resources Department to change the 401(k) allocation. Compared to the alternative activity, which has payoff 0, the investment activity has payoff $-k < 0$ at time t (the present) and payoff $b > 0$ for all periods from $t+1$ on. ($t+1$ included). The individual has to choose when to undertake the investment activity, that is, at t, at $t+1$, at $t+2$, etc. (The individual can also decide not to do it, which we define as doing at $t = \infty$) Assume that both k and b are deterministic.

a) Consider first a time-consistent individual ($\beta = \hat{\beta} = 1$) and solve for the optimal timing of the investment decision. **Show** that the optimal solution takes the form of a threshold rule as a function of k, δ, and b.

b) Consider then a sophisticated present-biased individual ($\beta = \hat{\beta} < 1$). Compute the utility for the current self from investing today, at time t. Compute the utility for the current self from investing T periods into the future, that is, at $t+T$.

c) Show that this implies that a sophisticated agent will wait for at most T days to invest if the cost of investing k satisfies

$$k \leq \frac{\beta\delta b}{1-\beta\delta^T}T \tag{2}$$

[You will need a Taylor expansion of $1-\delta^T$ for δ going to 1: $1-\delta^T \simeq (1-\delta)\,T$]

d) Consider now a fully naive present-biased individual ($\beta < \hat{\beta} = 1$). As of time t, under what conditions does the individual expect to invest tomorrow (at $t+1$)? Argue that the naive agent compares the utility from investing today and tomorrow.

e) Show that the fully naive present-biased individual invests at time t (and otherwise never invests) if and only if

$$k \leq \frac{\beta\delta b}{1-\beta\delta}.$$

f) In particular, discuss what the following sentence means: *The naive agent procrastinates if*

$$\frac{\beta\delta b}{1-\beta\delta} < k \leq \frac{\delta b}{1-\delta}.$$

What is the difference between procrastinating ($\frac{\beta\delta b}{1-\beta\delta} < k \leq \frac{\delta b}{1-\delta}$) and not investing ($k > \frac{\delta b}{1-\delta}$)? Explain intuitively.

g) Now we apply these calculations to address the evidence in Madrian and Shea (2001) and Choi et al. (2005). As in Question 2, consider a new employee in a company without automatic enrollment (that is, the default is no investment). On any day, the employee can pay an effort cost $k > 0$ and invest in the 401(k), thereafter reaping benefit b in every subsequent day. Can you provide reasonable values for k, δ and β for an individual with average earnings? Justify all the assumptions you make. (Remember: b and δ are on a *daily* scale).

h) In particular, which factors would enter into the determination of b? Can it be negative for some employees? Assume that it is positive in what follows.

i) Go back to the time-consistent employee in point (a). For what value of k should the individual be indifferent between investing and no (given the calibrated values of b, δ and β)? What do you expect the individual to do?

j) Move on now to the sophisticated present-biased employee in points (b) and (c). Using equation (2) calibrate the value of k for which the individual may wait 360 days (that is, one year) to invest, which is about the observed pattern. Are these plausible levels of the parameters, or do you expect that a sophisticated agent will invest earlier?

k) Consider now the naive present-biased employee in points (d)-(f). For realistic values of the parameters, is it likely that the employee will rationally delay ($k > \frac{\delta b}{1-\delta}$)? Is it likely that the employee will procrastinate ($\frac{\beta\delta b}{1-\beta\delta} < k \leq \frac{\delta b}{1-\delta}$)?

l) Which calibrated model fits the data better? Why? Now that you picked your favorite model, let's try to criticize that too. Which problems does your favorite model have?