This exam is comprised of three sections. The first section is for material covered in ECON 220A taught by Nano Barahona. The second section is for material covered in ECON 220B taught by Matthew Backus and Quitzé Valenzuela-Stookey. The third section is for material covered in ECON 220C taught by Ben Handel and Carolyn Stein. Each section has 100 points. You have to **answer only two sections**. If you attempt to answer more than two sections, be very clear about which two sections you want us to grade. We will only assign points on two sections. You have 3 hours to complete the exam.

1. Section #1 (220A) - 100 points

Miller and Weinberg (2017) (MW onwards) study the price effects of a merger between Miller and Coors in 2008. They use scanner data from the IRI Academic Database to estimate a model of supply and demand for beer.

They use a random coefficient nested logit (RCNL) model to estimate consumer demand. The indirect utility that consumer i receives from inside good j in region r and period t is

$$u_{ijrt} = x_j\beta_i^* + \alpha_i^* + \sigma_j^D + \tau_t^D + \xi_{jrt} + \bar{\varepsilon}_{ijrt}, \tag{1}$$

where x_j is a vector of observable characteristics, p_{jrt} is the retail price, σ_j^D allows the mean valuation of unobserved product characteristics to vary freely by product, σ_j^D allows the mean valuation of the indirect utility from consuming the inside goods to vary freely over time, ξ_{jrt} is an unobserved quality valuation specific to the region-period, and $\bar{\varepsilon}_{ijrt}$ is a stochastic term that follows the distributional assumptions of the nested logit model with one nest for the inside goods and one nest for the outside good.

The observable product characteristics include a constant (i.e., an indicator that equals 1 for an inside good), calories, package size, and an indicator for whether the product is imported. They control for σ_j^D and τ_t^D using product and time dummy variables, respectively, and specify the consumer-specific coefficients as $[\alpha_i^*, \beta_i^*]' = [\alpha, \beta]' + \Pi D_i$, where D_i is (demeaned) consumer income.

They use a model of differentiated-products price competition to estimate supply. The vector of equilibrium prices in each region-period satisfies the first-order condition

$$p_t = mc_t - \left[\Omega_t(\kappa) \circ \left(\frac{\partial s_t(p_t;\theta^D)}{\partial p_t}\right)^T\right]^{-1} s_t(p_t, \theta^D) \tag{2}$$

where Ω_t is the ownership matrix, s_t is a vector of market shares, and the operation *circ* is element-by-element matrix multiplication.

Answer the following questions about the paper. Be brief in your answers.

1. (10 points) Explain what is the main hypothesis that the authors are trying to test in the paper. How does it relate to Ω_t? Write down Ω_t as a function of κ before and after the merger.

2. (5 points) MW estimate demand using a random coefficient nested logit model. Write down the equation that determines the market shares as a function of the product characteristics, prices, and model parameters.

3. (21 points) To estimate the model, MW use the following instruments:

   (a) The distance between the brewery and the region interacted with diesel prices

   (b) The number of products in the market

   (c) The distance summed across all products in the market

   (d) Mean income interacted with observed product characteristics (a constant, calories, package size, and an import dummy)

   (e) An indicator equal to 1 for ABI and MillerCoors products after the merger

   (f) The number of products in the market interacted with indicators for ABI and Miller/-Coors products

   (g) The distance summed across all products in the market interacted with indicators for ABI and Miller/Coors products

   MW provide intuition about how each instrument "identifies" each of the non-linear parameters. Explain, for each set of instruments, what parameters help they identify most. Discuss the relevance condition and the exclusion restriction for each of them. Give one example of a violation of the exclusion restriction for each set.

4. (10 points) Explain what is the role of the fixed effects in the model and why are they necessary for the identification of α. Argue why including region fixed effects could have helped for the credibility of the identifying assumption via an example and discuss potential problems with that implementation.

5. (12 points) MW estimate the model following BLP (1995). Write down a pseudo-code that can estimate the model. Be explicit about every step.

6. (7 points) After estimating the model, MW recover a value of ρ of 0.8299 and Median Outside Diversion of 12.96%. Explain using words what the Median Outside Diversion is. If the estimated ρ was smaller, would the Median Outside Diversion be larger or smaller? Justify your answer.

7. (6 points) Explain what are the potential problems with the conclusions of the paper if MW had used a plain logit instead of a random coefficients nested logit to estimate demand.

8. (6 points) Explain three differences in model specification between MW's model and the model that Nevo (2001) uses to estimate demand for cereal. What are the pros on cons of their decisions in each dimension?

9. (10 points) Petrin (2002) introduces a new set of moments that can be used to identify the model parameters. Explain what kind of data would be needed to implement their method in the context of MW. Be explicit about the key data features that MW did not have access to that one would need and explain how that additional data would have helped to identify Π. Describe how to modify the algorithm from question 5 to incorporate that data.

10. (7 points) Moving on to the supply side, explain, in words, how is κ identified.

11. (6 points) Describe the main finding of the paper. What does this imply for antitrust merger analysis?

2. Section #2 (220B) - 100 points

Please note: all of these questions can be answered in just a few sentences or lines of algebra. If you get carried away writing long answers, you will risk running out of time.

2.1. Conduct Testing (25 points): In Backus et al. (2020) you learned about a procedure for testing models of conduct in firm pricing based on a comparison of the implied marginal costs. Suppose that you have already estimated demand $s_j(\mathbf{p})$ and the associated matrix of demand derivatives, $\Omega_{jk} = \partial s_j(\mathbf{p})/\partial p_k$.

BCS use a set of variables z to test the moment restriction $A(z_{jt}) \cdot \omega_{jt}^m = 0$ for competing models.

1. First, using the demand system, show how to solve for marginal costs under the behavioral assumption of "single-product pricing," i.e. that all products are priced as if they were owned by an independent firm.

2. For each of the following classes of variables, explain why it is either useful or not useful for discriminating between models of firm conduct in pricing:

   (a) Demand shifters.

   (b) Cost shifters.

   (c) BLP instruments.

3. Give an example of a pair of models of firm conduct in the pricing game that are *not* testable in this framework.

2.2. Production Function Estimation (25 points): We learned about production function estimation in the context of Ackerberg et al. (2015). For the purposes of this question, assume we are interested in a Cobb-Doubglas production technology with two inputs, k and ℓ, with k a dynamic input and ℓ perfectly variable:

$$y_{jt} = \beta^{\ell} \ell_{jt} + \beta^{k} k_{kt} + \omega_{jt} + \varepsilon_{jt}$$

Recall that ω_{jt} in this setting stands in for the part of productivity that is known to the firm but not the econometrician, while ε_{jt} is the part of productivity that is unknown to either.

1. Ackerberg et al. (2015) show that the labor coefficient β^{ℓ} is not identified in the first-stage regression of the Olley and Pakes (1996) framework, which used investment i_{jt} to construct a proxy for ω_{jt}

$$OP: \quad y_{jt} = \beta^{\ell} \ell_{jt} + \Phi(k_{jt}, i_{jt}) + \varepsilon_{jt}$$

   Briefly explain why β^{ℓ} is not identified.

2. Despite this, they still run their own version of first-stage regression, with ℓ_{jt} inside of $h(\cdot)$.

$$ACF: \quad y_{jt} = \Phi(k_{jt}, \ell_{jt}, i_{jt}) + \varepsilon_{jt}$$

   If this regression doesn't identify β^{ℓ}, what does it accomplish?

   * In case it is confusing, let me acknowledge that ACF use m_{jt} instead of i_{jt} as their proxy variable, following Levinsohn and Petrin (2003), but that is not important here.

3. Some have criticized Ackerberg et al. (2015), along with Olley and Pakes (1996) and Levinsohn and Petrin (2003), because the estimator does not permit generic models of market power in the output market. Give an example of how such market power could lead to a violation of the assumptions of the model.

2.3. Theory question (50 points): This question is based on Bergemann and Morris (2016). For clarity, the relevant definitions are included at the end of the section.

Consider a set of agents and a baseline information structure $S = (T, \pi)$. Imagine that the agents are firms in a market. These firms may be able to learn the private information of others (i.e. their types in the baseline information structure). Here we model this learning in a completely exogenous, non-strategic way: there is some *network*, N, such that if firms i and j are neighbors in this network then i observes t_j and j observes t_i. Let $\mathcal{N}$ be the set of all possible networks, which includes the complete network in which all firms' types are publicly observed.

There is a principal who wishes to impact the behavior of the firms (for this question it is not necessary to be more explicit about the principal's objective). The principal can do this by providing information: the principal can choose any expansion, S^*, of the baseline information structure S. As in Bergemann and Morris (2016), this is equivalent to saying that the principal can send messages to each firm, and the message sent can depend on the state and the realized type profile from the baseline information structure S.[1]

The challenge is that the principal has no information about the network, and wants to be robust to this uncertainty. The principal therefore plays the following game against nature:

1. The principal chooses the information structure S^*, which must be an expansion of S.
2. Nature chooses a network N from the set $\mathcal{N}$.
3. Firms' types are drawn according to the baseline information structure S. Firms then observe the types of their neighbors in the network chosen by nature. Then firms observe the messages sent by the principal and update their beliefs. (Note that they only observe their own message).
4. Firms play a BNE of the induced incomplete incomplete information game (if there are multiple BNE, assume they play the one that is best for the principal).

The objective of nature is to minimize the principal's payoff.

Question. Show that there is an optimal strategy for the principal in which the principal's messages consist of action recommendations to each firm, and these recommendations constitute a BCE regardless of which network nature chooses.

(HINT: Show that it is without loss of optimality for the principal to play as if nature has already chosen the complete network.)

(HINT: The more information a firm has, the more obedience constraints need to be satisfied in the definition of BCE.)

2.4. Definitions: *Agents' payoffs.* There are I agents, 1,2,..., I, with i as a typical agent. There is a finite set of states, Θ. Each agent has a finite set of actions, A_i, and we write $A := A_1 \times \cdots \times A_I$. Each agent has a utility function $u_i : A \times \Theta \to \mathbb{R}$.

Information. An information structure S consists of

i. For each agent i, a finite set of *types* (or signals) T_i, where $T := T_1 \times \cdots \times T_I$,

ii. A joint distribution of types and states, $\pi \in \Delta(T \times \Theta)$.

It is convenient to decompose $\pi(t, \theta) = \psi(\theta)\pi(t|\theta)$. A *decision rule* is a mapping $\sigma : T \times \Theta \to \Delta(A)$.

[1] To be clear, the principal always gets to observe each agent's type; while the network determines which agents observe each others' types.

Definition 1 (Obedience)**.** Decision rule σ is *obedient* given information structure S if for each $i = 1, \ldots, I$, $t_i \in T_i$, and $a_i \in A_i$, we have

$$\sum_{a_{-i}, t_{-i}, \theta} \psi(\theta)\pi(t|\theta)\sigma((a_i, a_{-i})|(t_i, t_{-i}), \theta) \left[u_i((a_i, a_{-i})|\theta) - u_i((a_i', a_{-i})|\theta)\right] \geq 0$$

for all $a_i' \in A_i$.

As in Bergemann and Morris (2016), a decision rule is a called a Bayes correlated equilibrium (BCE) if it is obedient. Bergemann and Morris (2016) show that the set of BCE is is the same as the set of Bayes-Nash equilibria (BNE) that can be induced by giving agents additional information about the state θ and type profile t.

A strategy for agent i is a function $\beta_i : T_i \to \Delta(A_i)$.

Definition 2 (Bayes-Nash equilibrium)**.** A strategy profile β is a Bayes-Nash equilibrium given information structure S if for each i, $t_i \in T_i$, and $a_i \in A_i$ with $\beta_i(a_i|t_i) > 0$, we have

$$\sum_{a_{-i}, t_{-i}, \theta} \psi(\theta)\pi(t|\theta) \left(\prod_{j \neq i} \beta_j(a_j|t_j)\right) \left[u_i((a_i, a_{-i})|\theta) - u_i((a_i', a_{-i})|\theta)\right] \geq 0$$

for all $a_i' \in A_i$.

Given two information structures $S^1 = (T^1, \pi^1)$ and $S^2 = (T^2, \pi^2)$, we say that the information structure $S^* = (T^*, \pi^*)$ is a *combination* of S^1 and S^2 if $T^* = T^1 \times T^2$ and the marginal distributions of of π^* on T^1 and T^2 are the same as under π^1 and π^2. Information structure S^* is an expansion of S if there is some information structure S' such that S^* is a combination of S and S'.

Theorem 1 (Bergemann and Morris (2016))**.** A decision rule σ is a Bayes correlated equilibrium given S if and only if, for some expansion S^* of S, there is a Bayes Nash equilibrium given S^* that induces σ.

3. Section #3 (220C) - 100 points

3.1. Adverse selection and inertia (25 points): This will be a multi-part question asking about selection markets

1. (10 points) In Handel (2013, Adverse Selection and Inertia) describe how inertia is separately identified from persistent unobserved preferences for insurance. Please do the following:
   (a) Write down the demand model in detail.
   (b) What is the main structural preference that is a persistent unobserved preference?
   (c) What features of the data allow those preferences to be separate from inertia?
   (d) How would additional insurance choices in a given year impact identification of those preferences and inertia?

2. (5 points) Now, imagine that the mechanism underlying inertia is not a "switching cost" but is instead some other micro-founded model for inertia, such as rational inattention or naive inattention (or some other foundation for inertia!). Write down a version of this alternative model that you could estimate, i.e. modify the model in part 1. to have this new micro-foundation for inertia.

3. (5 points) Describe in depth how you might empirically test whether your model in 2. is a better model than the switching cost model set up in Handel (2013).

4. (5 points) Describe in depth (i) if you think your model in 2. would have different implications for adverse selection than the switching cost model in the paper and, if so (ii) what would those implications be?

3.2. Insurance contracts (25 points):

1. (5 points) Describe the central tradeoff studied in HHW and what the authors find empirically regarding this tradeoff. Describe the tradeoff in the context of the degree of risk-rating the regulator allows in a competitive market.

2. (5 points) How does the regulatory / contract structure in Ghili et al. (2023) differ from that in HHW (2015)? How does the assumption relaxed in HHW (2015) for the Ghili et al. (2023) paper impact the tradeoff you discussed in 1.?

3. (5 points) For the contracts focused on in Ghili et al. (2023), how does steepness of lifetime income path impact welfare for those contracts vs. standard ACA exchange contracts with no risk rating allowed? Explain why using marginal utilities at different points in time as well as long run risk protection.

4. (5 points) The authors study model extensions that relate to (i) consumer myopia and (ii) switching costs. For each of these foundations describe (i) how contract structure changes and (ii) how welfare of using these contracts changes. Discuss as comparative static (i.e., what happens as switching costs or myopia increase / decrease).

5. (5 points) We don't see the contracts described in Ghili et al. (2023) much in practice. In addition to the economic fundamentals studied in the paper, what are some of the practical / logistical reasons the authors note for why we might not see those contracts much in practice?

3.3. Competition and Innovation (50 points):

1. [8 points] Theoretical models have been ambiguous as to whether competition increases or decreases innovation. The "Arrow effect" posits that competition should increase innovation. The "Schumpeter effect" posits that competition should decrease innovation. Explain both of these effects in 1-2 sentences each.

2. [8 points] In the paper "Competition and Innovation: An Inverted-U Relationship" Aghion et al. set up the following model:

   - There is a unit of mass of consumers who each provide one unit of labor inelastically.
   - Each consumer has the following preferences:

     $$u(y) = \ln y$$

     where y is made up of a continuum of intermediate goods:

     $$\ln y = \int_0^1 \ln x_j dj.$$

     This implies that consumers spend an equal share of income on all x_j. Normalize this common amount to 1.
   - The market for each good j is a duopoly between A and B. Consumers therefore maximize $x_j = x_{Aj} + x_{Bj}$ subject to $p_{Aj}x_{Aj} + p_{Bj}x_{Bj} = 1$.
   - The two firms have different production functions. For L units of labor, each firm can produce:

     $$x_{ij}(L) = \gamma^{k_i} L \text{ for } i = A, B$$

     where k indexes the level of technology in sector j and $\gamma > 1$.

   Suppose the wage for both firms is given exogenously by w. What are the costs for each firm of producing x_{Aj} and x_{Bj}, respectively?

3. [8 points] In a sector where firm A is technologically ahead (i.e., $k_A = k_B + 1$), what is firm A and firm B's profit?

4. [8 points] Let $\pi_{A,1}$ be firm A's profit in the problem above. In a sector where the firms are technologically even (i.e., $k_A = k_B$) assume that each firm will earn profits of $\pi_{A,0} = \pi_{B,0} = \varepsilon\pi_{A,1}$ where $\varepsilon \in [0, 1/2]$. Explain what the ε parameter measures, and why it must be between 0 and 1/2.

5. [10 points] Firms can try to increase k by a single unit. They must pay R&D costs to do so. The more they spend on R&D, the higher the probability that they increase k.

   (a) Consider firm B in the sector where firm A is technologically ahead, as in part 3. Suppose that ε decreases. Is firm B going to increase or decrease R&D spending? A qualitative argument is fine. Link this to either the Arrow or Schumpeter effect.

   (b) Consider firm B in the sector where the firms are technologically even, as in part 4. Suppose that ε decreases. Is firm B going to increase or decrease R&D spending? A qualitative argument is fine. Link this to either the Arrow or Schumpeter effect.

6. [8 points] In the paper "Killer Acquisitions," Cunningham, Ederer, and Ma show that large incumbent pharmaceutical firms sometimes acquire small firms working on projects that overlap with the incumbent firm. In many cases, the incumbents shut these projects down after acquiring them. Is this behavior an example of the Arrow effect or Schumpeter effect? Explain.