Spring 1997 John Rust

Economics 551b 37 Hillhouse, Rm. 27

FINAL EXAM

April 30, 1997

INSTRUCTIONS: Answer **3** of the questions below. Each question is worth 100 points, for a total of 300 possible points. You have 3 hours for the exam, or 1 hour per question.

QUESTION 1 (Hypothesis testing) Consider the GMM estimator with *IID* data, i.e the observations $\{y_i, x_i\}$ are independent and identically distributed. Show that in the overidentified case $(J > K)$ that the minimized value of the GMM criterion function is asymptotically χ^2 with $J - K$ degrees of freedom:

$$N H_N(\hat{\theta}_N)[\hat{\Omega}_N]^{-1} H_N(\hat{\theta}_N) \underset{d}{\Longrightarrow} \chi^2(J-K),$$

where H_N is a $J \times 1$ vector of moment conditions, θ is a $K \times 1$ vector of parameters, $\chi^2(J-K)$ is a Chi-squared random variable with $J - K$ degrees of freedom,

$$\hat{\theta}_N = \underset{\theta \in \Theta}{argmin}\, H_N(\theta)[\hat{\Omega}_N]^{-1} H_N(\theta),$$

$$H_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} h(y_i, x_i, \theta),$$

and $\hat{\Omega}_N$ is a consistent estimator of Ω given by

$$\Omega = E\{h(\tilde{y}, \tilde{x}, \theta^*) h(\tilde{y}, \tilde{x}, \theta^*)'\}.$$

QUESTION 2 (Markov Processes)

A. (10 points) Are Markov processes of any use in econometrics? Describe some examples of how Markov processes are used in econometrics such as providing models of serially dependent data, as a framework for establishing convergence of estimators and proving laws of large numbers, central limit theorems, etc. and as computational tool for doing simulations.

B. (10 points) What is a random walk? Is a random walk always a Markov process? If not, provide a counter-example.

C. (40 points) What is the ergodic or invariant distribution of a Markov process? Do all Markov processes have invariant distributions? If not, provide a counterexample of a Markov process that doesn't have an invariant distribution. Can a Markov process have more than 1 invariant distribution? If so, give an example.

D. (40 points) Consider the discrete Markov process $\{X_t\} = \{1, 2, 3\}$ with transition probability

$$\begin{aligned} P\{X_{t+1} = 1 | X_t = 1\} &= \frac{1}{2} \quad P\{X_{t+1} = 2 | X_t = 1\} = \frac{1}{3} \quad P\{X_{t+1} = 3 | X_t = 1\} = \frac{1}{6} \\ P\{X_{t+1} = 1 | X_t = 2\} &= \frac{3}{4} \quad P\{X_{t+1} = 3 | X_t = 2\} = \frac{1}{4} \quad P\{X_{t+1} = 2 | X_t = 3\} = 1 \end{aligned}$$

Does this process have an invariant distribution? If so, find all of them.

QUESTION 3 (Consistency of M-estimator) Consider an M-estimator defined by:

$$\widehat{\theta}_N = \arg\max_{\theta\in\Theta} Q_N(\theta).$$

Suppose following two conditions are given
(i) (Identification) For all $\varepsilon > 0$

$$Q(\theta^*) > \sup_{\theta\notin B(\theta^*,\varepsilon)} Q(\theta)$$

where $B(\theta^*,\varepsilon) = \{\theta \in R^k | \|\theta - \theta^*\| < \epsilon\}$.
(ii) (Uniform Convergence)

$$\sup_{\theta\in\Theta} |Q_N(\theta) - Q(\theta)| \xrightarrow{p} 0.$$

Prove consistency of the estimator by showing

$$P\left(\widehat{\theta}_N \notin B(\theta^*,\varepsilon)\right) \xrightarrow{p} 0.$$

QUESTION 4 (Consistency of Bayesian posterior) Consider a Bayesian who has observes *IID* data $(X_1,\ldots,X_N)$, where $f(x|\theta)$ is the likelihood for a single observation, and $p(\theta)$ is the prior density over an unknown finite-dimensional parameter $\theta \in R^K$.

A. (30 points) Let $P(\theta \in A|X_1,\ldots,X_N\}$ be the posterior probability θ is in some set $A \subset \Theta$ given the first N observations. Show that this posterior probability satisfies the *Law of iterated expectations:*

$$E\{P(\theta \in A|X_1,\ldots,X_{N+1})|X_1,\ldots,X_N\} = P(\theta \in A|X_1,\ldots,X_N).$$

B. (20 points) A *martingale* is a stochastic process $\{\tilde{Z}_t\}$ that satisfies $E\left\{\tilde{Z}_{t+1}|\mathcal{I}_t\right\} = \tilde{Z}_t$, where $\mathcal{I}_t$ denotes the information set at time t and includes knowledge of all past Z_t's up to time t, $\mathcal{I}_t \supset (\tilde{Z}_1,\ldots,\tilde{Z}_t)$. Use the result in part A to show that the process $\{\tilde{Z}_t\}$ where $\tilde{Z}_t = P(\theta \in A|\tilde{X}_1\ldots,X_t)$ is a martingale. (We are interested in martingales because the *Martingale Convergence Theorem* can be used to show that if θ is finite-dimensional, then the posterior distribution converges with probability 1 to a point mass on the true value of θ generating the observations $\{X_i\}$. But you don't have to know anything about this to answer this question.)

C. (50 points) Suppose that if θ is restricted to the K-dimensional simplex, $\theta = (\theta_1,\ldots,\theta_K)$ with $\theta_i \in (0,1)$, $i = 1,\ldots,K$, $1 = \sum_{i=1}^K \theta_i$, and the distribution of X_i given θ is multinomial with parameter θ, i.e.

$$Pr\{X_i = k\} = \theta_k, \quad k = 1,\ldots,K.$$

Suppose the prior distribution over θ, $p(\theta)$ is *Dirichlet* with parameter α:

$$p(\theta) = \frac{\Gamma(\alpha_1+\cdots+\alpha_K)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_K)}\theta_1^{\alpha_1-1}\cdots\theta_K^{\alpha_K-1}$$

where both $\theta_i > 0$ and $\alpha_i > 0$, $i = 1,\ldots,K$. Compute the posterior distribution and show 1) the posterior is also Dirichlet (i.e. the Dirichlet is a conjugate family), and show directly that as $N \to \infty$ that the posterior distribution converges to a point mass on the true parameter θ generating the data.

QUESTION 5 (Time series question) Suppose $\{X_t\}$ is an ARMA(p,q) process, i.e.

$$A(L)X_t = B(L)\epsilon_t$$

where $A(L)$ is a q^{th} order lag-polynomial

$$A(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q$$

and $B(L)$ is a p^{th} order lag-polynomial

$$B(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots + \beta_p L^p$$

and the lag-operator L^k is defined by

$$L^k X_t = X_{t-k}$$

and $\{\epsilon_t\}$ is a white-noise process, $E\{\epsilon_t\} = 0$ and (cov(ϵ_t, ϵ_s)= 0 if $t \neq s$, $= \sigma^2$ if $t = s$).

A. (30 points) Write down the autocovariance and spectral density functions for this process.

B. (30 points) Show that if $p = 0$ an autoregression of X_t on q lags of itself provides a consistent estimate of $(\alpha_0/\sigma, \ldots, \alpha_q/\sigma)$. Is the autoregression still consistent if $p > 0$?

C. (40 points) Assume that a central limit theorem holds, i.e. the distribution of normalized sums of $\{X_t\}$ to converge in distribution to a normal random variable. Write down an expression for the variance of the limiting normal distribution.

QUESTION 6 (Empirical question) Assume that shoppers always choose only a single brand of canned tuna fish from the available selection of K alternative brands of tuna fish each time they go shopping at a supermarket. Assume initially that the (true) probability that the decision-maker chooses brand k is the same for everybody and is given by θ^*_k, $k = 1, \ldots, K$. Marketing researchers would like to learn more about these choice probabilities, $\theta^* = (\theta^*_1, \ldots, \theta^*_K)$ and spend a great deal of money sampling shoppers' actual tuna fish choices in order to try to estimate these probabilities. Suppose the Chicken of the Sea Tuna company undertook a survey of N shoppers and for each shopper shopping at a particular supermarket with a fixed set of K brands of tuna fish, recorded the brand b_i chosen by shopper i, $i = 1, \ldots, N$. Thus, $b_1 = 2$ denotes the observation that consumer 1 chose tuna brand 2, and $b_4 = K$ denotes the observation that consumer 4 chose tuna brand K, etc.

A. (10 points) Without doing any estimation, are there any general restrictions that you can place on the $K \times 1$ parameter vector θ^*?

B. (10 points) Is it reasonable to suppose that θ^*_k is the same for everyone? Describe several factors that could lead different people to have different probabilities of purchasing different brands of tuna. If you were a consultant to Chicken of the Sea, what additional data would you recommend that they collect in order to better predict the probabilities that consumers buy various brands of tuna? Describe how you would use this data once it was collected.

C. (20 points) Using the observations $\{b_1, \ldots, b_N\}$ on the observed brand choices of the sample of N shoppers, write down an estimator for θ^* (under the assumption that the "true" brand choice probabilities θ^* are the same for everyone). Is your estimator unbiased?

D. (20 points) What is the maximum likelihood estimator of θ^*? Is the maximum likelihood estimator unbiased?

E. (40 points) Suppose Chicken of the Sea Tuna company also collected data on the prices $\{p_1, \ldots, p_K\}$ that the supermarket charged for each of the K different brands of tuna fish. Suppose someone proposed that the probability of buying brand j was a function of the prices of all the various brands of tuna, $\theta_j^*(p_1, \ldots, p_K)$, given by:

$$\theta_j^*(p_1, \ldots, p_K) = \frac{\exp\{\beta_j + \alpha p_j\}}{\sum_{k=1}^K \exp\{\beta_k + \alpha p_k\}}$$

Describe in general terms how to estimate the parameters $(\alpha, \beta_1, \ldots, \beta_K)$. If $\alpha > 0$, does an increase in p_j decrease or increase the probability that a shopper would buy brand j?

QUESTION 7 (Regression question) Let (y_t, x_t) be *IID* observations from a regression model

$$y_t = \beta x_t + \epsilon_t$$

where y_t, x_t, and ϵ_t are all scalars. Suppose that ϵ_t is normally distributed with $E\{\epsilon_t|x_t\} = 0$, but $\text{var}(\epsilon_t|x_t) = \sigma^2|x_t|^\theta$. Consider the following two estimators for β^*:

$$\hat{\beta}_T^1 = \frac{\sum_{t=1}^T y_t}{\sum_{t=1}^T x_t}$$

$$\hat{\beta}_T^2 = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2}$$

A. (20 points) Are these two estimators consistent estimators of β^*? Which estimator is more efficient when: 1) if we know *a priori* that $\theta^* = 0$, and 2) we don't know θ^*? Explain your reasoning for full credit.

B. (20 points) Write down an asymptotically optimal estimator for β^* if we know the value of θ^* *a priori.*

C. (20 points) Write down an asymptotically optimal estimator for (β^*, θ^*) if we don't know the value of θ^* *a priori.*

D. (20 points) Describe the feasible GLS estimator for (β^*, θ^*). Is the feasible GLS estimator asymptotically efficient?

E. (20 points) How would your answers to parts A to D change if you didn't know the distribution of ϵ_t was normal?