Economics 696F, Final Exam

This take-home final exam is due on Wednesday, May 3, by 5 pm, in my office. (You may email the exam to me if you prefer, but please send it as a PDF file, not a Word document.)

For the final exam, you must work completely on your own. You may use any books, lecture notes, and other printed reference materials you wish, but you may not discuss your exam with any other person while you are working on it.

1. Consider a nonparametric regression model with additive separability: for y_i, x_i, z_i all scalar random variables,
$$E[y_i|x_i, z_i] = m_1(x_i) + m_2(z_i).$$
Suggest a nonparametric estimator for $m_1(\cdot)$ and $m_2(\cdot)$. Explain it in enough detail so that a skilled programmer who did not know statistics or econometrics could implement it. Be explicit about how to select any "smoothing parameters" (e.g. bandwidths, truncation values).

2. Suppose that $x_i, i = 1, \ldots, n$ are iid $Unif[\theta, \theta + 1]$.
   (a) Calculate $E_\theta[x_i]$ and use this to suggest a GMM estimator. What is the asymptotic distribution of the GMM estimator? (Give as much detail as possible.)
   (b) Write the likelihood function. Is the maximum likelihood estimator well-defined in the sense that there is a unique solution to the problem of maximizing the likelihood?
   (c) Show that the MLE is consistent and derive its asymptotic distribution. (If the MLE is not unique, you will need to define some unique version of the MLE, and then derive its asymptotic properties.) Compare the MLE and GMM estimators.

3. Suppose we have two random samples. Random sample 1 is a sample of size n_1 from the joint distribution of (y, x). In other words, we observe (y_i, x_i) iid from some joint density $f(y, x)$, for $i = 1, \ldots, n_1$.
Random sample 2 is a sample of size n_2 from the marginal distribution of y. So we also draw y_i iid from $f(y) = \int f(y, x)dx$, for $i = n_1 + 1, \ldots, n_1 + n_2$.
Assume that
$$E[y|x] = x'\beta.$$
   (a) Sketch a GMM estimator for β that uses both sample 1 and 2. Do we expect an efficiency gain compared to using only sample 1 and estimating β by OLS?

(b) Now suppose that in sample 1, y is interval censored. We do not observe y, but we observe v_0 and v_1, and we assume that $Pr(v_0 \leq y \leq v_1) = 1$. In sample 2, y is not censored in any way.

   i. Explain how to estimate β using only sample 1.

   ii. Can the data in sample 2 be used to improve the estimate of β? If so, explain how.

4. (Based on Chu, Leslie, and Sorensen, 2006.) We want to model consumer choice over two goods, which are made available separately or as a "bundle" which is priced differently from the prices of the individual items. We can think of the consumer choosing one of the following bundles:

   1. item 1 only
   2. item 2 only
   3. bundle of items 1 and 2
   4. nothing

   We assume that consumer i $(= 1, \ldots, n)$ has latent utility for bundle j $(= 1, \ldots, 4)$ given by:

$$u_{ij} = \begin{cases} V_{i1} - \alpha \cdot p_{ij} & \text{for } j = 1 \\ V_{i2} - \alpha \cdot p_{ij} & \text{for } j = 2 \\ V_{i1} + V_{i2} - \alpha \cdot p_{ij} - \beta_i & \text{for } j = 3 \\ 0 & \text{for } j = 4. \end{cases}$$

   Here, p_{ij} is the price for bundle j faced by consumer i, and we assume that V_{i1}, V_{i2}, and β_i have a joint multivariate normal distribution with mean vector μ and variance matrix Σ. The price coefficient α is assumed to be strictly positive. (In principle, we could allow α to vary across individuals, but for this exercise treat it as fixed.)

   We observe for each individual the vector of prices $p_i = (p_{i1}, p_{i2}, p_{i3})$, and which bundle they chose (1 through 4).

   Derive a simulation based estimator for α, μ, Σ. Provide pseudo-code, or enough detail so that a skilled programmer who was not familiar with this type of model could implement the estimator.