# FINAL EXAM

The exam has 6 questions worth a total of 200 points. It should not take more than about 4 hours.

(1) (50 points) Suppose

$$y_i = x_i'\beta + z_i\theta + \varepsilon_i,\ i = 1, \ldots, n$$

where $\{y_i, x_i, z_i\}_{i=1}^n$ are observed, $\varepsilon_i \sim$iid $\mathcal{N}(0,1)$, $x_i$ is a $k \times 1$ nonstochastic regressor, and $z_i$ is a scalar nonstochastic regressor. We are interested in testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$. We want the test to be invariant to the transformations

$$\{y_i, x_i, z_i\}_{i=1}^n \to \{y_i + x_i'b, x_i, z_i\}_{i=1}^n,\ b \in \mathbb{R}^k. \tag{1}$$

(a) Show that the transformations (1) form a group.
(b) Find a maximal invariant.
(c) Suppose $n > k$. Find the form of the uniformly most powerful invariant test.
(d) What is the best invariant test if $n \leq k$?

(2) (20 points) Suppose we observe the iid variables $y_i$, $i = 1, \ldots, n$ that are generated by

$$y_i = S_i Z_i + (1 - S_i)(\mu + Z_i)$$

where $S_i$ is iid Bernoulli($p$), and $Z_i$ is iid $\mathcal{N}(0,1)$ independent of $\{S_i\}$ (so the density of $y_i$ is a mixture between $\mathcal{N}(0,1)$ and $\mathcal{N}(\mu,1)$, with mixing weights $p$ and $1-p$). Suppose we are interesting in testing $H_0 : \mu = 0$.

(a) Do you think that a textbook MLE-based test will have the usual properties in large samples?
(b) Derive the form of the test that maximizes weighted average power, with a weighting function that specifies $\mu \sim \mathcal{N}(0, \omega^2)$ and $p$ uniform $[0,1]$. Provide a reaonably simplified expression for the test statistic.

(3) (30 points) Suppose $y_t = \mu + u_t$, $u_t = \rho u_{t-1t} + \varepsilon_t$, where $\varepsilon_t \sim iid(0, \sigma^2)$, $|\rho| < 1$ and $\mu$ is a scalar parameter.

(a) What is the long-run variance $\omega^2$ of $y_t$?
(b) Consider a standard non-parametric estimator $\hat{\omega}^2$ of the long-run variance $\omega^2$ with a particular kernel, and let $b_T$ be its mean square error minimizing bandwidth. How does $b_T$ vary qualitatively with $\rho$? And what can you say about the sign of the bias $E[\hat{\omega}^2 - \omega^2]$?

(4) (50 points) Consider the system

$$x_t = 1.45x_{t-1} - .55y_{t-1} + \varepsilon_{xt} \tag{2}$$

$$y_t = .45x_{t-1} + .45y_{t-1} + \varepsilon_{yt}\ , \tag{3}$$

where $\varepsilon_x$ and $\varepsilon_y$ are the innovations of $x$ and $y$ in this bivariate system and both have variance 1.

(a) Show that the system is cointegrated and display it in VECM form.

(b) Show that the differenced data vector, $(\Delta x_t, \Delta y_t)$ does not have an autoregressive representation.

(c) Suppose that, noting that both series seem to be non-stationary, an econometrician worked with the differenced data and fit just a second-order bivariate VAR to $\Delta x$ and $\Delta y$. Calculate the covariance matrix of residuals from this model and compare it to that of the true first-order model in levels. [Hint: Though the differenced data have no AR representation, they are stationary and have an MA representation, from which the covarinaces at all leads and lags can be determined. But your computations probably have to be approximate and done with the computer. The frequency domain and back to find the autocovariance function of the differenced data, followed by cholesky decomposition or regression projection to find the finite-order AR might be the best route.]

(5) (50 points) One can take a Bayesian approach to the mixture model of problem (2).

(a) Describe how, using a $N(0, \omega^2)$ prior for $\mu$ and a uniform prior on $p$, and drawing $\{S_i\}$ sequences as well as values for $p$ and $\mu$, one could use pure Gibbs sampling to generate a sample from the joint posterior on $p$ and $\mu$, as well as posterior density values at the points drawn. Explain exactly what distribution should be drawn from at each stage of the sampling scheme.

(b) Show how the marginal data density (integal of prior times likelihood) can be computed analytically for the model with the restriction $\mu = 0$.

(c) Describe an algorithm that could use the posterior density values from the Gibbs sampling to compute the marginal data density for the model with $\mu$ unconstrained.

(d) Gelman in *Bayesian Data Analysis* argues that it is almost always possible to find a better approach than comparing models with posterior odds. Is this example such a case? Why or why not?