Problem Set # 5

Econ 722

1. Calculate the KL divergence from a $N(\mu_0, \sigma_0^2)$ distribution to a $N(\mu_1, \sigma_1^2)$ distribution.

2. Suppose we observe a random sample $\{(\mathbf{x}_t, y_t)\}_{t=1}^T$ from some population and decide to forecast y from $\mathbf{x}$ using the following linear model:

$$y_t = \mathbf{x}_t'\beta + \varepsilon_t$$

Let $\widehat{\beta}$ denote the ordinary least squares estimator of β based on $\{(\mathbf{x}_t, y_t)\}_{t=1}^T$. Now suppose that we observe a *second* random sample $\{(\widetilde{\mathbf{x}}_t, \widetilde{y}_t)\}_{t=1}^T$ from the sample population that is *independent* of the first. Show that

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\widehat{\beta})^2\right] \leq \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}(\widetilde{y}_t - \widetilde{\mathbf{x}}_t'\widehat{\beta})^2\right]$$

In other words, show that the in-sample squared prediction error is an overly optimistic estimator of the out-of-sample squared prediction error.

3. This question asks you to fill in part of the derivation of AIC$_c$ for a normal regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_0^2\mathbf{I}_T)$ and k is the number of regressors. Define

$$\widehat{KL} = \frac{T}{2}\left[\frac{\sigma_0^2}{\widehat{\sigma}^2} - \log\left(\frac{\sigma_0^2}{\widehat{\sigma}^2}\right) - 1\right] + \left(\frac{1}{2\widehat{\sigma}^2}\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$

where $(\widehat{\beta}, \widehat{\sigma}^2)$ are the maximum likelihood estimators of (β_0, σ_0^2). Show that

$$\mathbb{E}\left[\widehat{KL}\right] = \frac{T}{2}\left\{\frac{T+k}{T-k-2} - \log(\sigma_0^2) + \mathbb{E}[\log\widehat{\sigma}^2] - 1\right\}$$

4. This question is based on Hurvich & Tsai (1993).
   (a) Write a short summary (2–3 paragraphs) of the paper and its main findings.
   (b) Replicate the simulation experiment from Section 4 of the paper, but with 5000 rather than 100 simulation replications. Briefly discuss your results.