Detecting Stage-Wise Outliers in Hierarchical Bayesian Linear Models of Repeated Measures Data

Mario Peruggia[1], Thomas J. Santner[1], and Yu-Yun Ho[2]

[1] The Ohio State University
Department of Statistics
1958 Neil Avenue
Columbus, OH 43210-1247

[2] Novartis Pharmaceuticals Corporation
Biostatistics & Statistical Reporting
One Health Plaza
Eastr hanover, NJ 07936-1080

July 2003

Running Title: Hierarchical Bayes Outlier Detection

Keywords: Autoregressive errors; BUGS; Graphical diagnostics; Model-based diagnostics; Outlier accommodation models; Diagnostics for multi-stage models.

Abstract

We propose numerical and graphical methods for outlier detection in hierarchical Bayes modeling and analyses of repeated measures regression data from multiple subjects; data from a single subject are generically called a "curve." The first-stage of our model has curve-specific regression coefficients with possibly autoregressive errors of a prespecified order. The first-stage regression vectors for different curves are linked in a second-stage modeling step, possibly involving additional regression variables. Detection of the *stage* at which the curve appears to be an outlier and the *magnitude and specific component* of the violation at that stage is accomplished by embedding the null model into a larger parametric model that can accommodate such unusual observations.

As a first diagnostic, we examine the posterior probabilities of first-stage and second-stage anomalies relative to the modeling assumptions for each curve. For curves where there is evidence of a model violation at either stage, we propose additional numerical and graphical diagnostics. For first-stage violations, the diagnostics identify the specific measurements within a curve that are anomalous. For second-stage violations, the diagnostics identify the curve parameters that are unusual relative to the pattern of parameter values for the majority of the curves. We give two examples to illustrate the diagnostics, develop a BUGS program to compute them using MCMC techniques, and examine the sensitivity of the conclusions to the prior modeling assumptions.