Journal of Educational and Behavioral Statistics
Winter 2001, Vol. 26, No. 4, pp. 443–468

Detecting a Change in School Performance: A Bayesian Analysis for a Multilevel Join Point Problem

Yeow Meng Thum
University of California, Los Angeles

Suman K. Bhattacharya
R. W. Johnson Pharmaceutical Research Institute

A substantial literature on switches in linear regression functions considers situations in which the regression function is discontinuous at an unknown value of the regressor, X_k*, where* k *is the so-called unknown "change point." The regression model is thus a two-phase composite of* $y_i \sim N(\beta_{01} + \beta_{11}x_i, \sigma_1^2)$*,* $i = 1, 2, \ldots, k$ *and* $y_i \sim N(\beta_{02} + \beta_{12}x_i, \sigma_2^2)$*,* $i = k + 1, k + 2, \ldots, n$*. Solutions to this single series problem are considerably more complex when we consider a wrinkle frequently encountered in evaluation studies of system interventions, in that a system typically comprises multiple members (*$j = 1, 2, \ldots, m$*) and that members of the system cannot all be expected to change synchronously. For example, schools differ not only in* whether *a program, implemented system-wide, improves their students' test scores, but depending on the resources already in place, schools may also differ in* when *they start to show effects of the program. If ignored, heterogeneity among schools in when the program takes initial effect undermines any program evaluation that assumes that change points are known and that they are the same for all schools. To describe individual behavior within a system better, and using a sample of longitudinal test scores from a large urban school system, we consider hierarchical Bayes estimation of a multilevel linear regression model in which each individual regression slope of test score on time switches at some unknown point in time,* k_j*. We further explore additional results employing models that accommodate case weights and shorter time series.*

Keywords: *change and join point, hierarchical Bayes, longitudinal data, Markov Chain Monte Carlo, multilevel modeling, piecewise regression, program evaluation, school performance*

To evaluate the effect of a program on a certain relevant measure of school performance, the educational researcher could compare the school's performance after the intervention with its performance before. Frequently, the researcher compares the postintervention mean on a standardized test with its preintervention mean. A

This research was supported in part by the Center for Research on Evaluation, Standards, and Student Testing (CRESST). A grant to the senior author from the Council on Research of the Academic Senate of the Los Angeles Division of the University of California provided additional support. In addition to helpful suggestions from the associate editor and three anonymous referees, the authors appreciate comments from Tony Bryk, Bengt Muthén, and Michael Seltzer on an earlier version of this article.

better gauge of the program effects on performance can be obtained, if repeated measurements are available, by comparing the postintervention and preintervention trends in a piecewise regression of performance measure on time. This practice, however, assumes that the time of intervention, t, coincides with the point in time, k, at which the program takes initial effect. Although a clear improvement, the analysis may be misleading if the change point, k, is in fact unknown and different from t.

Figure 1 illustrates what can go wrong with the usual piecewise regression for this situation if the assumption that change in school is coincident with the intervention point is mistaken. Suppose we denote the preintervention and postintervention slopes as $\beta_1^{(t)}$ and $\beta_2^{(t)}$, respectively, if we assume that change occurred at time t, and let $\beta_1^{(k)}$ and $\beta_2^{(k)}$ denote the pre and postintervention slopes, respectively, if change had occurred at k. Figure 1a depicts the situation in which an intervention at time t is coincident with when change starts, k. An evaluation based on this assumption correctly estimates the change in slope, as $(\beta_2^{(t)} - \beta_1^{(t)}) = (\beta_2^{(k)} - \beta_1^{(k)})$. This same analysis would however underestimate the effect if change actually begins at $k > t$, as depicted by the dashed lines in Figure 1b, because we suspect that $(\beta_2^{(k)} - \beta_1^{(k)}) \geq (\beta_2^{(t)} - \beta_1^{(t)})$. Figure 1c suggests considerably greater confusion for a routine multisite evaluation when change points, k_j, varies with site (such as schools, indexed by j) and are asynchronous with the time of intervention, t.

The literature on switching linear regression functions considers typical situations in which the regression function is discontinuous at an unknown value of the regressor, X_k, where k is the change point. The regression model is thus the two-phase composite

$$y_i \sim N\left(\beta_{0p} + \beta_{1p}x_i, \sigma_p^2\right), \tag{1}$$

where $i = 1, 2, \ldots, n$, $p = 1$ if $i \leq k$ and $p = 2$ if $i > k$.

Following Quandt (1958), similar attempts to reflect the uncertainty in change points in two-phase linear regression analysis have since appeared. The literature for two-phase regression is enormous, but a brief overview may be organized along three related themes. The first reveals a shared concern across various empirical research domains in identifying and detecting change in the course of developmental processes. Many applications are found in econometrics. Brown, Durbin, and Evans (1975) provide instances involving changes over time in the number of local telephone calls, in the demand for money, and in staff requirements in an organization. In climatology, Maronna and Yohai (1978) examine annual precipitation over time for change. In geology, Esterby and el-Shaarawi (1981) employ a two-phase polynomial to describe change in measures of pollen concentration in lake sediment cores obtained at various depths. Morrell, Pearson, Carter, and Brant (1995), Slate and Cronin (1997), and Slate and Clark (1999) presented nonlinear regression models with transition smoothing functions at the unknown change point to monitor changes in Prostate-Specific Antigen (PSA) profiles as a means for early prostate cancer detection. In epidemiology, Joseph, Wolfson, du Berger, and Lyle (1996) are concerned that a prepost comparison may be biased if the inter-

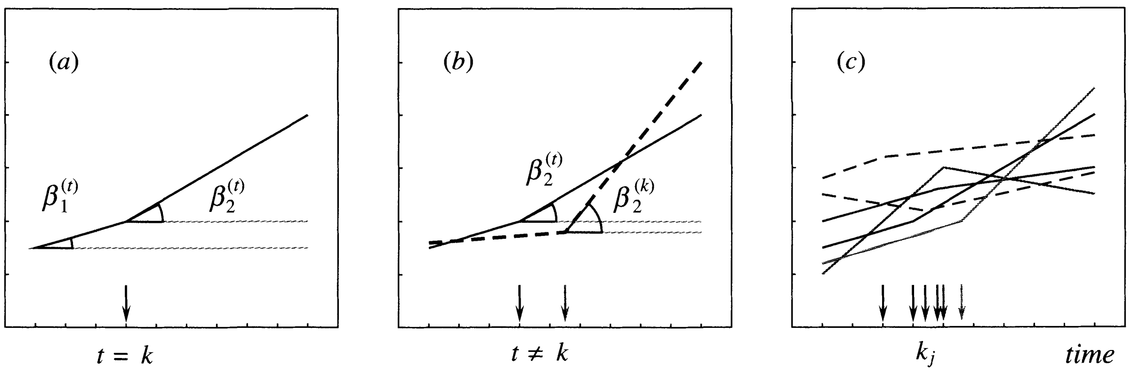

FIGURE 1. *Comparing post and preintervention regression slopes when (a) change for a school occurs at a known time point* k *which is coincident with the time of intervention* t, *(b) change point* k *is neither known nor coincident with* t *(c) schools change asynchronously and their change points are unknown.*

vention point is mistaken for the change point in their study on the effects of dietary calcium supplementation on high blood pressure.

A second theme in the research literature dwells on variants of Quandt's original formulation of the switching regression function, Equation 1, itself: whether the regression segments share a common intercept (a join point problem, e.g., Bacon & Watts, 1971), share a common slope but display a shift in their means (a mean shift problem, e.g., Hinkley & Schechtman, 1987), and share the same residual error variance (e.g., Worsley, 1983). Picard (1985) provides a more general consideration of unknown change points in time series analyses. Finally, the literature may also be organized along more methodological lines, with authors employing maximum likelihood solutions (e.g., Jandhyala & Fotopoupos, 1999), Bayesian methods (e.g., El-Sayyad, 1975), random regression mixtures (e.g., Quandt & Ramsey, 1978), as well as nonparametric approaches (e.g., Wolfe & Schechtman, 1984). The interested reader is directed to the comprehensive reviews of Hinkley, Chapman, and Runger (1980) and Zacks (1983). More recent efforts, laced with a stronger Bayesian flavor, extend beyond the two-phase normal linear regression to other developmental processes. Müller and Rosner (1994) study triphasic linear models using a semiparametric Bayesian approach. Raftery and Akman (1986) and Carlin, Gelfand, and Smith (1992) formulate Bayesian procedures for changes in Poisson processes for count data, while Stephens (1994), Slate and Cronin (1997), and also Chib (1998) considered problems with more than one change point.

The solution to Quandt's single series problem, Equation 1, is considerably more complex when we consider a wrinkle frequently encountered in evaluation studies of system interventions in that a system typically comprises multiple members ($j = 1, 2, \ldots, m$) and that, furthermore, members of the system cannot all be expected to behave similarly, or otherwise change synchronously.

For a commonplace example in educational research, consider the putative effects of a large-scale intervention on student academic performance. Figure 2 shows the variability of school means for third-grade Iowa Tests of Basic Skills (ITBS) mathematics scores for a sample from Chicago Public Schools from 1988 to 1996. (Years are labeled 1 through 9 in the sequel.) For this analysis, we have placed the criterion-referenced test scores on an arbitrary linear scale. Displaying a series of box-plots for school test-score means over time invites inappropriate analyses which assume that school change is synchronous. The evidence suggests that schools vary in their patterns of change, a fact better represented by a plot of raw school test score profiles, as in Figure 3. Here, according to one interpretation, schools appear to differ not only in *whether* a program, implemented system-wide, improves their students' test scores, but depending on the resources already in place, schools may also differ in *when* they start to experience effects of the program. If ignored, heterogeneity among schools in when the program "kicks in" undermines any program evaluation that assumes that change points are known and that they are the same for all schools. It is important to recognize that an explanation of school reform in terms of the changes in test scores is not the goal

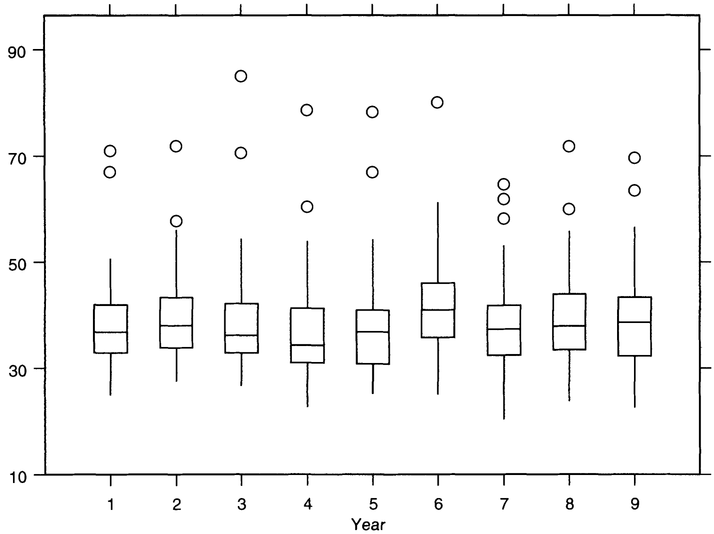

FIGURE 2. *Distributions of school Grade 3 ITBS math means.*

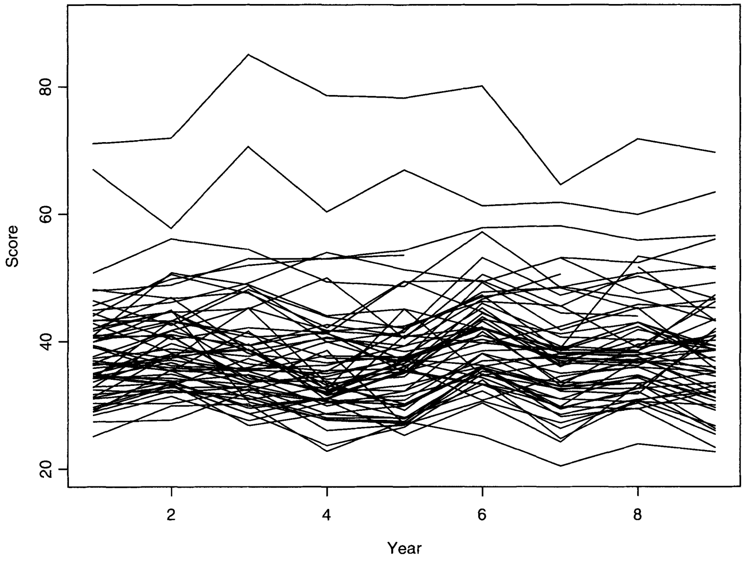

FIGURE 3. *Observed school Grade 3 ITBS math profiles.*

of the analyses. Any direct relation would certainly be naive given that many other unspecified causal mechanisms may also be at play in this context. Nevertheless, the issues considered here are theoretically instructive because how we determine the timing of change is critical to evaluation efforts for understanding what works in schools.

We consider a fully parametric hierarchical Bayesian estimation of a multilevel linear regression model in which each individual regression slope of test score on time changes at some unknown change point, X_{kj} unique to each school, j. Our approach and rationale closely resembles the multipath change point analysis in Joseph et al. (1996). They consider randomized trials in which the blood pressure of individuals under the same experimental conditions is not expected to respond to dietary calcium supplementation in the same way, nor within the same time frame. They suggest that a sound analysis must also account for the mediating effects of individual metabolism, as may be evidenced by variation in individual times taken in response to treatment. However, we extend their mean-change model by (a) estimating join point regression models for each individual school and, because the number of time points is relatively small and the within-school variability appears considerable, we also (b) reformulate the school level model with t errors at the school level. Also, because we expect that the uncertainty of a join point estimate becomes more considerable the shorter the time series, we also showed how inferences on school change itself can be easily constructed from the conditional posterior of the change in slopes, $(\beta_{2j} - \beta_{1j} \mid k_j = \hat{k}_j)$, where $\hat{k}_j$ is the modal estimator of the join point $_{kj}$, for example.

Our basic model is also similar to another recent study by Slate and Clark (1999), which traces the change in a bio-marker to give an early detection for prostate cancer for individual patients. In their application join points vary among units, but are assumed to be continuous rather than discrete. Both of the studies above share the major goals of our general modeling framework, which is to reflect individual differences in development within a system better when the timing of change is unknown.

Our study contributes to the literature on change point analysis for studying a bundled system of change processes. In the context of programmed interventions in school systems, the analysis brings an increased measure of sensitivity to program evaluation. It could, for example, help us answer the question of whether the overall improvement in performance could have resulted from a well-publicized intervention, given that improvement for some schools began later than projected. Additionally, this model is easily extended to accommodate school and community characteristics as covariates at the school level, an analytic strategy that could help identify and explain a school's delay in showing the anticipated effects of an intervention.

In the next section, we provide an overview of Quandt's change point model, and describe the features of the hierarchical Bayes formulation due to Carlin et al. (1992). We detail extensions to the multilevel change point regression assuming normally distributed errors. We then examine an analysis using data from an ongoing study of school trends in test scores conducted by the Consortium on Chicago School Research. An extension to our basic approach will be suggested with illustrative analyses incorporating case-weights. With another extension, we further evaluated our results for sensitivity to outlying observations through the use of t distributed errors. We conclude based on our preliminary evidence that join points for individual school Grade 3 ITBS mathematics profiles (from 1988–1996) indeed differ among a sample of urban elementary schools. The estimated posterior distribution of the join points suggests that although the estimated timings of change in Grade 3 mathematics performance do not contradict the claim that school reform may have been a contributing factor, changes have nonetheless not been uniformly positive.

Single Series Solutions

Suppose we observe multiple test performance profiles for a sample of schools in a system. A single series change point solution would model each series separately.

Maximum Likelihood

For Equation 1, Quandt (1958) shows that the log likelihood for fixed k is proportional to

$$-k \log \hat{\sigma}_1 - (n - k) \log \hat{\sigma}_2 .$$

That k is not continuous suggests that we take the maximum likelihood estimate of k to be the value of k that corresponds to the maximum *maximorum.* The likelihood ratio test against the full hypothesis

$$y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

is $\ell = \max_k \ell(k)$. Here,

$$\ell(k) = n \log \hat{\sigma}^2 - k \log \hat{\sigma}_1^2 - (n - k) \log \hat{\sigma}_2^2,$$

$k = 3, 4, \ldots, n - 3$, and $\hat{\sigma}_p^2$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of σ_p^2 and σ^2, respectively. Further details of this model and its subsequent development, including tests of a related model that assumes equal variances, are given by Worsley (1983).

Hierarchical Bayes

Carlin et al. (1992) pose Equation 1 on page 444 as the first in a three-stage hierarchical Bayes linear regression model. At the second stage of this model $\boldsymbol{\beta}_1 = (\beta_{01}, \beta_{11})'$ and $\boldsymbol{\beta}_2 = (\beta_{02}, \beta_{12})'$ are independent $N(\boldsymbol{\gamma}, \mathbf{T})$ where $\mathbf{T}$ is 4×4. σ_1^2 and σ_2^2 are independent $IG(a_0, b_0)$. A discrete uniform, U_n, represents our prior knowledge of the unknown change point k. Stage three hyperpriors in this model for $(\gamma, \mathbf{T})$ are normal–Wishart; $\gamma \sim N(\mu, \mathbf{C})$; and $\mathbf{T} \sim Inv\text{–}W(\mathbf{S}^{-1}, \rho)$.

The intermediate objective for the Gibbs solution is to derive the marginal posterior of k. Standard results from the multivariate normal show that the conditional posterior for each regression segment is,

$$\boldsymbol{\beta}_p \sim N\left(\mathbf{V}_p^k \mathbf{b}_p^k, \mathbf{V}_p^k\right),$$
$$\mathbf{V}_p^k = \left(\sigma_p^{-2} \mathbf{X}_p^{k\prime} \mathbf{X}_p^k + \mathbf{T}^{-1}\right)^{-1},$$
$$\mathbf{b}_p^k = \left(\sigma_p^{-2} \mathbf{X}_p^{k\prime} \mathbf{y}_p^k + \mathbf{T}^{-1}\boldsymbol{\gamma}\right),$$
$$\mathbf{y}_1^k = (y_1, \ldots, y_k)', \qquad \mathbf{y}_2^k = (y_{k+1}, \ldots, y_n)',$$
$$\mathbf{X}_1^k = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_k \end{pmatrix}' \text{ and } \mathbf{X}_2^k = \begin{pmatrix} 1 & \cdots & 1 \\ x_{k+1} & \cdots & x_n \end{pmatrix}'$$

Furthermore, the full conditional distributions of the unknowns (σ_1^2, σ_2^2, γ, $\mathbf{T}$, $\boldsymbol{k}$) can be given as

$$\sigma_1^2 \sim IG\left\{a_0 + \frac{k}{2}, \left[\frac{1}{2}\left(\mathbf{y}_1^k - \mathbf{X}_1^k \boldsymbol{\beta}_1\right)'\left(\mathbf{y}_1^k - \mathbf{X}_1^k \boldsymbol{\beta}_1\right) + b_0\right]\right\},$$

$$\sigma_2^2 \sim IG\left\{a_0 + \frac{n-k}{2}, \left[\frac{1}{2}\left(\mathbf{y}_2^k - \mathbf{X}_2^k \beta_2\right)'\left(\mathbf{y}_2^k - \mathbf{X}_2^k \beta_2\right) + b_0\right]\right\},$$
$$\boldsymbol{\gamma} \sim N\left\{\boldsymbol{\Delta}\left[\mathbf{T}^{-1}(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2) + \mathbf{C}^{-1}\mu\right], \boldsymbol{\Delta}\right\},$$
$$\mathbf{T} \sim inv - W\left\{\left[\sum\nolimits_p (\boldsymbol{\beta}_p - \boldsymbol{\gamma})(\boldsymbol{\beta}_p - \boldsymbol{\gamma})' + \mathbf{S}\right]^{-1}, \rho + 2\right\}.$$

Delta, ($\boldsymbol{\Delta}$), the variance-covariance matrix of the full conditional distribution for γ, is $(2\mathbf{T}^{-1} + \mathbf{C}^{-1})^{-1}$. The full conditional distribution for the join point, k, is in turn

$$p\left(k \mid \mathbf{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2, \boldsymbol{\gamma}, \mathbf{T}\right) = \frac{L\left(\mathbf{y}; k, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2\right)}{\sum_{U_n} L\left(\mathbf{y}; k, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2\right)}$$

where the likelihood is

$$L\left(\mathbf{y}; k, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma_1^2, \sigma_2^2\right) = \frac{\exp\left[-\frac{1}{2}\sum_p \left(\mathbf{y}_p^k - \mathbf{X}_p^k \boldsymbol{\beta}_p\right)' \left(\mathbf{y}_p^k - \mathbf{X}_p^k \boldsymbol{\beta}_p\right)/\sigma_p^2\right]}{\sigma_1^k \sigma_2^{n-k}}.$$

Single Data Series Example

Before we proceed with the case of multiple time series, we compare results for the single series formulations above for a simulated data series with the evaluation of change based on piecewise regression. Without loss of generality, we would work with a join point regression model (Cohen & Kushary, 1994) denoted as follows:

$$y_i \sim N\left[\beta_0 + \beta_1 \min(0, x_i - x_k) + \beta_2 \max(0, x_1 - x_k), \sigma^2\right]. \tag{2}$$

If $2 < k < (n - 1)$ for example, the predictor matrix $\mathbf{X}^k$ is

$$\mathbf{X}^k = \begin{pmatrix} 1 & x_1 - x_k & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{k-1} - x_k & 0 \\ 1 & 0 & 0 \\ 1 & 0 & x_{k+1} - x_k \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_n - x_k \end{pmatrix}.$$

We argue that, for shorter time series with no dramatic level change expected, a model such as Equation 2, with constant error variance for which only the slope changes after a join point, appears realistic. The first coefficient, β_0, is the expected value of the outcome variable at the join point, k. β_1 represents the regression slope before and up until the join point, and β_2 is the slope thereafter. Alternative codings for $\mathbf{X}^k$ are of course possible, including a parameterization which estimates directly the difference, $(\beta_2 - \beta_1)$, representing a change in slopes. For our illustration, we generated the series

$$y_i = (3.98, 3.38, 3.41, 3.33, 2.75, 3.10, 3.19, 2.96, 3.03, 2.94)$$

for $i = 1, 2, \ldots, 10$ based on model 2 above, setting $n = 10$, $k = 5$, $\beta_0 = 3.0$, $\beta_1 = -.2$, $(\beta_2 - \beta_1) = .2$, and $\sigma = .15$.

Results for ordinary least squares regression in Table 1 show that, not surprisingly, the model is misspecified if we are mistaken about when change actually occurred. If we are wrong about when change actually occurred, we fail to detect a positive change in regression slopes. The maximum likelihood solution correctly identifies $k = 5$ for this series, with regression estimates $\hat{\beta}_0 = 2.981$, $\hat{\beta}_1 = -.213$, and $\hat{\beta}_2 = 0.009$. Table 2 gives the solution, based on 10,000 updates, for Carlin's hierarchical Bayes approach, employing the discrete prior,

$$U_n = (.0, .05, .14, .14, .14, .14, .14, .13, .12, .0),$$

for the unknown change point. With only 10 observations, our choice for U_n is motivated by our presumption that the regimes before and after the change point are both linear, so that it seems reasonable to force k from the extremes of the series at the very least. This is ultimately a model selection issue that is beyond the scope of this article.

The point estimates in Table 2 are of limited use for inference, however, because they average over a change point distribution in U_n. A critical feature for the Gibbs

TABLE 1
OLS Piecewise Regression Results for Simulated Data Series.

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_2-\hat{\beta}_1$	R^2
No change	3.249*	−.085*			.561
	(.078)	(.027)			
Change at k					
3	3.219*	−.336*	−.040	.296	.742
	.123	.115	.030	.133	
4	3.104*	−.248*	−.022	.226	.752
	.121	.073	.035	.097	
5	2.980*	−.213*	.009	.222*	.800
	.109	.048	.038	.077	
6	2.956*	−.158*	.014	.172	.704
	.132	.046	.059	.093	

Note. *(Prob > $|t|$) ≤ 0.05.

solution, indeed a significant advantage, is the ability to closely examine the marginal posterior distribution of k, in Figure 4, for symmetry and multimodality. Figure 4 suggests that the mode, at $k = 5$, probably summarizes the marginal distribution more adequately, in agreement with our previous solution via maximum likelihood. The conditional posterior means for the regression function given $k = 5$ are provided in the lower portion of Table 2. These conditional results are comparable to the previous ordinary least squares and maximum likelihood solutions for change occurring at time point 5, with a 0.988 probability that the change in slopes is positive, that is, $p\ (\beta_2 \geq \beta_1 \mid k = 5)$.

TABLE 2
Some Features of the Marginal and Conditional Posterior Distributions for Simulated Data.

Parameter	*M*	*SD*	25%	*Mdn*	95%
Features of the Marginal Posteriors					
β_0	3.100	0.210	2.713	3.085	3.499
β_1	−0.301	0.221	−0.901	−0.240	−0.045
β_2	−0.016	0.099	−0.148	−0.023	0.160
$\beta_2-\beta_1$	0.287	0.236	−0.102	0.250	0.872
σ	0.219	0.077	0.123	0.203	0.412
k	4.365	1.849	2	4	9
Posterior Features Conditional on Join Point Mode, $k = 5$					
β_0	2.982	0.130	2.729	2.982	3.240
β_1	−0.213	0.057	−0.327	−0.214	−0.099
β_2	0.008	0.045	−0.083	0.008	0.094
$\beta_2-\beta_1$	0.221	0.092	0.038	0.222	0.400
σ	0.195	0.061	0.115	0.185	0.339

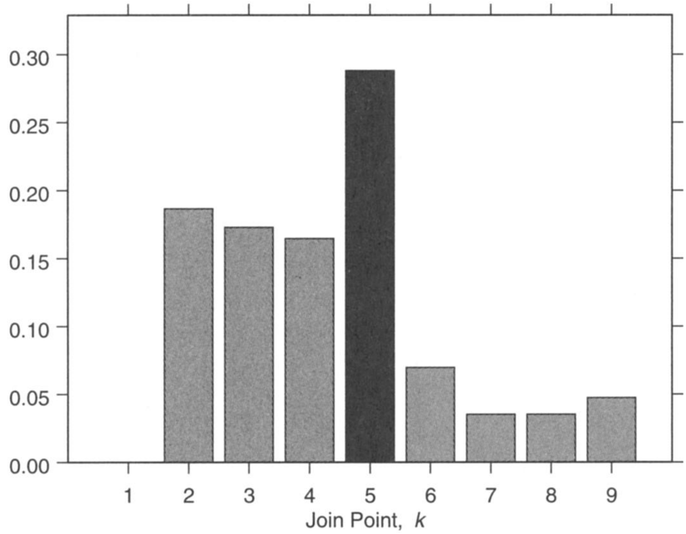

FIGURE 4. *Marginal posterior distribution of join point for simulated data series.*

Multilevel Regression with Random Join Points

If the essential features of each data series are considered exchangeable, the researcher will also be interested in characterizing parameters of the population. We derive our multilevel regression model with random change point guided by earlier results from Carlin et al. (1992) and Joseph et al. (1996).

The Model

For a sample of schools, the multilevel formulation for model 2 assumes $\boldsymbol{\beta}_j = (\beta_{0j}, \beta_{1j}, \beta_{2j})'$ are independent $N(\boldsymbol{\gamma}, \mathbf{T})$ and σ^2 is distributed $IG(a, b)$. Hyperpriors in this model for $(\gamma, \mathbf{T}^{-1})$ take the normal-Wishart form as before. The discrete uniform, $U(\pi_1, \pi_2, \ldots, \pi_n)$, represents our prior knowledge of the unknown join point k_j, and $\boldsymbol{\pi}'$ is distributed as a Dirichlet $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Results resemble closely those of Carlin et al. (1992). The conditional posterior of β_j is

$$\boldsymbol{\beta}_j \sim N\left(\mathbf{V}_j^{k_j}\mathbf{b}_j^{k_j}, \mathbf{V}_j^{k_j}\right),$$

where

$$\mathbf{V}_j^{k_j} = \left(\sigma^{-2}\mathbf{X}_j^{k_j\prime}\mathbf{X}_j^{k_j} + \mathbf{T}^{-1}\right)^{-1},$$

$$\mathbf{b}_j^{k_j} = \left(\sigma^{-2}\mathbf{X}_j^{k_j\prime}\mathbf{y}_j^{k_j} + \mathbf{T}^{-1}\gamma\right).$$

Implementing the Gibbs Sampler

From the above specification, the joint distribution of the data and all parameters is proportional to

$$\prod_j p(\mathbf{y}_j \mid k_j, \boldsymbol{\beta}_j, \sigma^2) \cdot p(\boldsymbol{\beta}_j \mid \gamma, \mathbf{T}) \cdot p(\gamma \mid \boldsymbol{\mu}, \mathbf{C}) \cdot p(\mathbf{T} \mid \mathbf{S}, \rho) \cdot p(\sigma^2 \mid a, b) \cdot p(k_j \mid \boldsymbol{\pi}) \cdot p(\boldsymbol{\pi} \mid \alpha)$$

To implement the Gibbs sampler, we require the full conditional distributions for $(\sigma^2, \boldsymbol{\gamma}, \mathbf{T}, \boldsymbol{\pi})$:

$$\sigma^2 \sim IG\left\{m\left(a + \frac{n}{2} + 1\right) - 1, \left[\frac{1}{2}\sum_j^m \left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j}\beta_j\right)'\left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j}\beta_j\right) + bm\right]\right\},$$

$$\boldsymbol{\gamma} \sim N\left[\Delta\left(\mathbf{T}^{-1}\sum_j \boldsymbol{\beta}_j + \mathbf{C}^{-1}\boldsymbol{\mu}\right), \Delta\right],$$

$$\mathbf{T} \sim Inv\text{-}W\left\{\left[\sum_j^m (\boldsymbol{\beta} - \gamma)(\boldsymbol{\beta} - \gamma) + \mathbf{S}\right]^{-1}, \rho + m\right\} .$$

Δ, the variance of the full conditional distribution for γ, is $(m\mathbf{T}^{-1} + \mathbf{C}^{-1})^{-1}$. The conditional distribution for the join point, k_j, is in turn

$$p(k_j = i \mid \mathbf{y}_j, \boldsymbol{\beta}_j, \sigma^2) = \frac{L(\mathbf{y}_j; k_j, \beta_j, \sigma^2) \cdot \pi_i}{\sum_i^n L(\mathbf{y}_j; k_j, \beta_j, \sigma^2) \cdot \pi_i},$$

where the likelihood is

$$L(\mathbf{y}_j; k_j, \beta_j, \sigma^2) = \exp\left[-\left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j}\beta_j\right)'\left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j}\beta_j\right)\Big/2\sigma^2\right]\Big/\sigma^n .$$

The discrete uniform prior for join point, k_j, may be represented as

$$p(k_j \mid \boldsymbol{\pi}) = \pi_1^{I_1(k_j)}\pi_2^{I_2(k_j)}\ldots\pi_n^{I_n(k_j)} ,$$

by using the indicator function

$$I_i(k_j) = \begin{cases} 1 & \text{if } k_j = i, \\ 0 & \text{otherwise.} \end{cases}$$

The Dirichlet ($\boldsymbol{\alpha}'$) hyperprior, the conjugate prior for the probabilities π' (DeGroot, 1970), is

$$p(\boldsymbol{\pi} \mid \boldsymbol{\alpha}) = \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \pi_i^{\alpha_i - 1}.$$

Conditional on (k_j, α), π is distributed

$$\begin{aligned} p(\boldsymbol{\pi} \mid \mathbf{k}, \boldsymbol{\alpha}) &\propto \prod_j^m \prod_i^n \left(\pi_i^{I_i(k_i)} \pi_i^{\alpha_i - 1}\right) \\ &= \prod_i^n \pi_i^{\sum_j^m I_i(k_j)} \pi_i^{m\alpha_i - m} \\ &= \prod_i^n \pi_i^{\left\{m(\alpha_i - 1) + \sum_j^m I_i(k_j)\right\}}, \end{aligned}$$

which is Dirichlet $[m(\alpha_i - 1) + \sum_j^m I_i(k_j)]$, so that the full conditional for π' is

$$\boldsymbol{\pi} \propto \exp\left[-\sum_j^m \left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j} \boldsymbol{\beta}_j\right)' \left(\mathbf{y}_j^{k_j} - \mathbf{X}_j^{k_j} \boldsymbol{\beta}_j\right)/2\sigma^2\right] \times \prod_i^n \pi_i^{m(\alpha_i - 1) + \sum_j^m I_i(k_j)}.$$

Estimation of the parameters of interest requires iterative Monte Carlo integration. Following Gelfand and Smith (1990), we perform the integration using Markovian updating via the Gibbs sampler.

Academic Outcomes and School Reform

Recent research on the academic productivity of Chicago's public elementary schools concludes that there was system-wide improvement in Grade 3 mathematics learning, as measured with the ITBS, from 1987 through 1996 (Bryk, Thum, Easton, & Luppescu, 1998). Bryk's three-level hierarchical linear regression models an individual student's input to the grade and the gain he makes in that school. That is, the first stage student-level model employs both the student's Grade 3 test score (output from Grade 3) as well as his Grade 2 test score (input to Grade 3), along with their individual standard errors of measurement. Data are longitudinal within the school. Presuming growth is linear throughout, trends for input and for gain over time are estimated for each school. These growth factors are then allowed to vary across schools in the system.[1]

A natural follow-up question, in a politically sensitive school reform environment, is whether the observed improvement is the result of school reform. Specifically, do the gains occur within some reasonable time frame after the Chicago School Reform Act of 1988? Although suggestive of a positive school reform effect to the advocate, this is not a question the original analysis is set up to answer and none is ventured. First, if reform has had an effect it is believed not to have been appreciable until at least 1990. The value of t for 1990 under model 2 is 3, two years after the legislation has passed, when it is argued that school resources and reorganization were finally in place for most schools in the system. This phenomenon cannot be captured by linear growth parameterization with a unitary slope used in Bryk et al.'s stage two model. Instead, a two-phase regression on time with the break-point at 1990 would be necessary, a strategy that nevertheless also depends on the unstated assumption of synchronous change, occurring in 1990. This assumption appears unlikely from Figure 3. We attempt to give a tentative answer to this question, showing how we may evaluate the impact of system-wide school reform employing our multilevel join point analysis, Equation 2, using a representative subset of the schools ($m = 58$). If the reform is causal of positive changes in academic performance, we expect to see school test score trends change for the better *after* 1990. In the present analysis, however, student gains are not the focus. We use only school means calculated from students who have been in a same school for two consecutive assessments. For simplicity, analyses involving student and school covariates will be considered elsewhere.

Model Hyperpriors

We employed the following conjugate hyperpriors in our multilevel Bayesian join point regression analysis:

$$a = 0 \qquad b = 1000,$$

$$\boldsymbol{\mu}' = (0, 0, 0) \qquad \mathbf{C} = \begin{pmatrix} 10^4 & 0 & 0 \\ 0 & 10^4 & 0 \\ 0 & 0 & 10^4 \end{pmatrix},$$

$$\mathbf{S} = \begin{pmatrix} .5 & 0 & 0 \\ 0 & .25 & 0 \\ 0 & 0 & .5 \end{pmatrix} \qquad \rho = 4,$$

$$\boldsymbol{\alpha}' = (\tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}).$$

Unlike the single series example above, we did not constrain k_j away from each end of the time period. Our experience indicates that with multiple series, there is a borrowing of information when we treat k_j as exchangeable, leading to shrinkage in the distribution of k_j as reflected by the marginal posterior of $\boldsymbol{\pi}$. In our analyses, we also experimented with alternative noninformative priors, especially with the Dirichlet ($\boldsymbol{\alpha}'$) because they are the principal objects of our inference. In general, we observe substantial differences in convergence rates but reasonably comparable marginal estimates. All calculations are obtained using the Gibbs sampler implemented in BUGS (Spiegelhalter, Thomas, Best, & Gilks, 1995). Diagnostics suggest that the solution, based on updates totaling 30,000, converged.

Results

Table 3 summarizes results from the multilevel join point analysis. Figure 5, in particular, plots the marginal posteriors for $\boldsymbol{\pi}_i$, and shows that only for $i = k = 3$ is there density appreciably higher than the equally likely prior probability of $1/9$. Thus, pooling information across schools in the multilevel analysis, also an attractive feature for Joseph et al. (1996), suggests that most of the school regressions switched

TABLE 3
Multilevel Join Point Solution: Marginal Posterior Features.

Hyperparameter	*M*	*SD*	25%	*Mdn*	95%
Regression Parameters					
γ_0	39.000	1.347	36.370	38.990	41.650
γ_1	0.174	0.255	−0.335	0.173	0.673
γ_2	0.127	0.213	−0.310	0.130	0.540
Variance Components of Regression Parameters					
τ_{11}	90.120	18.860	59.890	87.870	133.500
τ_{12}	7.743	3.037	2.599	7.441	14.620
τ_{13}	−2.872	1.661	−6.526	−2.743	0.034
τ_{22}	1.186	0.525	0.466	1.087	2.481
τ_{23}	−0.258	0.204	−0.719	−0.241	0.092
τ_{33}	0.465	0.187	0.203	0.432	0.918
Error Variance					
σ^2	3.299	0.120	3.074	3.295	3.542
Posterior Probability at Join Points					
π_1	0.092	0.082	0.002	0.068	0.304
π_2	0.089	0.082	0.002	0.065	0.304
π_3	0.194	0.125	0.015	0.176	0.478
π_4	0.129	0.111	0.004	0.099	0.414
π_5	0.105	0.093	0.003	0.078	0.344
π_6	0.095	0.085	0.003	0.071	0.316
π_7	0.106	0.094	0.003	0.080	0.346
π_8	0.096	0.089	0.003	0.070	0.327
π_9	0.093	0.083	0.003	0.070	0.308

at $k_j = 3$, that is, in 1990, which may be good evidence for attributing school improvement to school reform (lacking other competing explanations of course).

If we fix the join point for a school at the modal value of join point, k_j, we obtain the fitted piecewise school trends in Figure 6. We base our inference on the growth factors for the school on the posterior distributions of β_{0j}, β_{1j}, and β_{2j} conditional on the modal estimate of k_j because, although it does not reflect completely the uncertainty in k_j, its determination is based not just on the data for a school but from a pooling of information from schools in the population. Employing the marginal posterior distributions in this case will over emphasize the uncertainty in determining k_j; but that may sometimes appear preferable (see Joseph et al., 1996).

Our results further indicate that schools have not uniformly improved. Table 4 shows the means of conditional posterior distributions of the regression functions for each school for the subset of schools with change occurring away from each extreme of the time span. The largest slope gain is [.151 − (−1.592)] = 1.643 score units per year, showing a productivity gain of some 1.643 × 3 ≈ 5 score points from 1993 through 1996 (School 54). About 14 schools improve with change coming at the heels of reform in 1990 or shortly thereafter, for $p(\beta_{2j} \geq \beta_{1j}) \mid k_j$ greater than .80. After

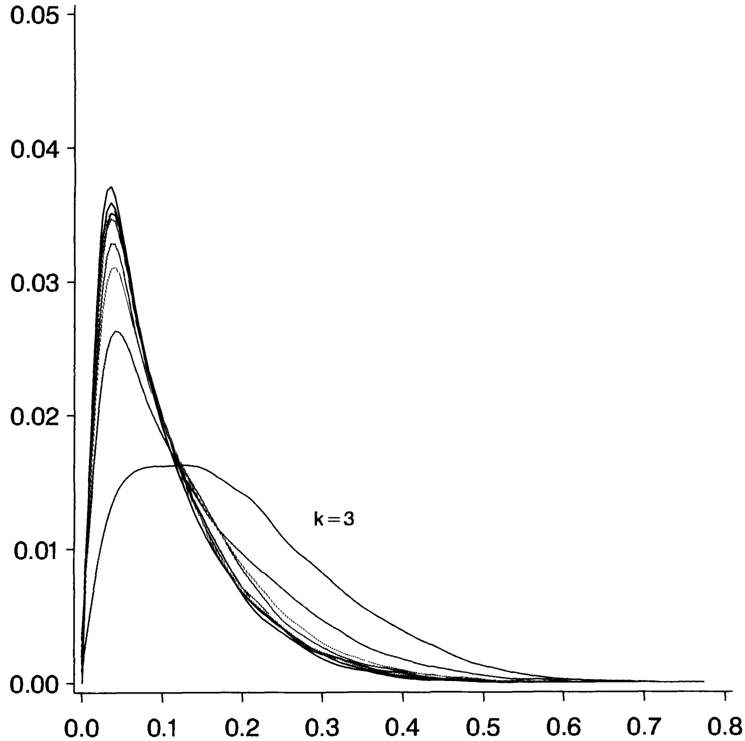

FIGURE 5. π', *Marginal posterior distribution of the probability of join point at* k.

1990 changes in their slopes are positive, of at least 1.1 score units per year each. This analysis also suggests that an almost equal number of schools show declines after 1990, with slope changes of at least −1.0 and with probability greater than 0.8.

Figure 7a shows a scatterplot of the posterior means of change in slopes, $(\hat{\beta}_{2j} - \hat{\beta}_{1j})$, versus the slopes before the detected join point, $\hat{\beta}_{1j}$. The size of the plot symbol varies proportionally with the expected attainment level, $\hat{\beta}_{0j}$, at the join point. The analysis shows that schools that have performed relatively well, for example, schools with higher estimated join points such as School 17 and School 45, generally take a turn for the worse. On the other hand, among poorly performing schools (e.g., School 54), changes in slopes are on the whole positive. Based on some initial analyses to be considered elsewhere, we suspect that the strongly negative correlation between the growth rate prior to change and the change in growth rates afterwards is quite typical of piecewise models for developmental processes over the shorter time frame.

Alternative Models

Two characteristics of our application deserve further attention: (a) the school means we have employed are measured with varying precision, and (b) trend esti-

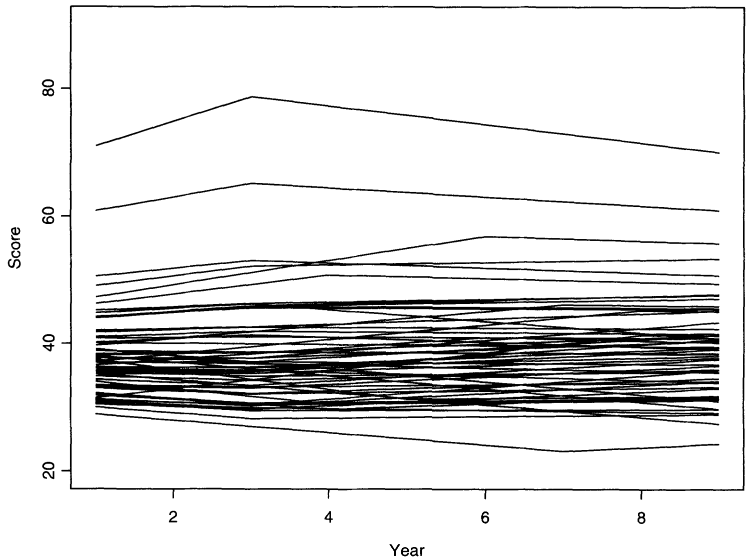

FIGURE 6. *Estimated school trends for Modal* k_j.

mates for shorter time series data can be especially sensitive to influential or outlying observations. Both factors present a potential danger to a routine regression analysis. However, they also provide a good opportunity to present simple extensions to our basic approach.

Case Weighting

Recall that our data comprise annual summaries in the form of grade-level test score means. Because schools not only differ from one another in the number of third grade classes they offer, the number of third grade classrooms within a school may also vary over time. At the same time, enrollment often fluctuates from year to year within a classroom. The result is that school-grade means are typically measured with varying degrees of precision. Under these circumstances, we can strengthen our previous exploratory study of our school test score data considerably by weighting the means we have for each year in each school by the number of observations, $n_{ij}\, r_{ij}$, on which the means are based. The weighted analysis begins with Equation 2. We simply multiply the ith row of $[\mathbf{y}_j\, \mathbf{X}_j^k]$ by $\sqrt{n_{ij}}$, and proceed with the previously outlined Gibbs sampler.

Shorter Time Series

As far as trend estimation is concerned, nine observations might be considered barely adequate with noisy data although many studies in the social sciences have touted results based on trend estimates with as few as three or four repeated observations.[2] We explore a minor extension to our multilevel random join point model above for

TABLE 4
Mean Estimates of School Regressions Conditional on Modal Join Point.

School	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$	$p(\lVert\beta_{2j} \geq \beta_{1j}\rVert \mid k_j)$
				.5————.8——.9—
17	78.666	3.865	−1.435	: . . . . : . . . . : . . 3
45	65.152	2.175	−0.708	: . . . . : . . . . : . . 3
14	56.953	1.939	−0.393	: . . . . : . . . . : . . 6
51	50.766	1.521	−0.268	: . . . . : . . . . : . . 4
20	52.975	1.243	−0.378	: . . . . : . . . . : . 3 :
10	46.223	0.716	−0.905	: . . . . : . . . . : . 3 :
33	52.104	1.549	0.207	: . . . . : . . . 3 . . :
57	46.072	1.071	−0.109	: . . . . : . . . . : 7 . :
1	41.012	0.887	−0.275	: . . . . : . . . 6 . . :
30	41.376	0.605	−0.369	: . . . . : . 3 . : . . :
43	45.795	0.935	−0.036	: . . . . : . 3 . : . . :
53	45.509	0.689	−0.024	: . . . . 3: . . . . : . . :
11	45.838	0.847	0.202	: . . . . 3 . : . . . . : . . :
48	45.866	0.886	0.308	: . . . . 3 . . : . . . . : . . :
56	41.888	0.402	−0.062	: . . . . 3 . . . : . . . . : . . :
34	42.558	0.367	−0.028	: . . . . 3 . . . . : . . . . : . . :
41	46.268	0.541	0.224	: . . . .3. . . . . . : . . . . : . . :
16	38.680	0.168	−0.055	: . . .3. . . . . . . : . . . . : . . :
32	35.788	−0.140	−0.297	: . . 3 . . . . : . . . . : . . :
55	41.020	0.089	−0.036	: . 3 . . . . : . . . . : . . :
37	36.850	−0.116	−0.234	: . 3 . . . . : . . . . : . . :
3	39.906	−0.081	−0.085	3 . . . . : . . . . : . . :
39	38.572	−0.128	−0.016	: 3. . . . . : . . . . : . . :
40	42.617	0.275	0.412	: . 3 . . . . : . . . . : . . :
44	29.855	−0.274	−0.120	: . . 3 . . . . : . . . . : . . :
21	33.195	−0.157	0.124	: . . . . 3 . . . . : . . . . : . . :
28	38.538	0.176	0.462	: . . . . 3 . . . . : . . . . : . . :
29	37.180	−0.363	−0.077	: . . . . 3 . . . . : . . . . : . . :
24	36.254	−0.038	0.300	: . . . . . 3 . . . . : . . . . : . . :
46	35.307	0.034	0.421	: . . . . . . 3 . . . . : . . . . : . . :
23	34.838	−0.352	0.130	: . . . . 3 . . . . . . : . . . . : . . :
12	36.241	0.053	0.538	: . . . . 3 . . . . . . : . . . . : . . :
15	37.699	−0.038	0.468	: . . . . 3 . . . . . . : . . . . : . . :
35	35.571	−0.002	0.592	: . . . . 3 . . . . : . . . . : . . :
4	34.203	−0.360	0.301	: . . . . 3 . . . : . . . . : . . :
27	32.034	−0.493	0.199	: . . . . 3 . . . : . . . . : . . :
22	29.421	−0.657	0.047	: . . . . 3 . . . : . . . . : . . :
36	37.090	0.020	0.730	: . . . . 3 . . . : . . . . : . . :
2	30.652	−0.618	0.112	: . . . . 3 . . . : . . . . : . . :
49	32.978	−0.150	0.592	: . . . . 3 . . . . . . : . . :
19	35.670	−0.010	0.840	: . . . . : 3 . . : . . :
7	30.260	−0.719	0.207	: . . . . : . 3 . : . . :

(continued)

TABLE 4 (*Continued*)
Mean Estimates of School Regressions Conditional on Modal Join Point.

School	$\hat{\beta}_{0j}$	$\hat{\beta}_{1j}$	$\hat{\beta}_{2j}$	$p(\lVert\beta_{2j} \geq \beta_{1j}\rVert \mid k_j)$
				.5————.8—.9—
5	37.719	−0.003	0.937	: . . . . : . 3 . : . . :
18	30.339	−0.895	0.143	: . . . . : . 3 . : . . :
8	32.165	−0.636	0.403	: . . . . : . . 3 : . . :
31	28.169	−0.922	0.126	: . . . . : . 3 . : . . :
6	30.248	−0.511	0.646	: . . . . : . . . 3 . . :
42	33.455	−0.463	0.701	: . . . . : . . . 3 . . :
47	29.666	−0.704	0.553	: . . . . : . . . 3 . . :
50	30.353	−0.441	0.923	: . . . . : . . . : 3 . :
25	30.462	−1.208	0.225	: . . . . : . . . : 4 . :
13	23.052	−0.979	0.638	: . . . . : . . . : . . 7
54	31.185	−1.592	0.151	: . . . . : . . . : . . 5 .

Note. Symbols for $p(\lVert\beta_{2j} \geq \beta_{1j}\rVert \mid k_j)$ in last column signify modal join points.

our data by replacing the assumption of normally distributed errors with a heavy-tailed density such as the t distribution. With degrees of freedom λ set smallish, at about 4, the t robustifies inferences against moderate misspecification of the distributional assumption when the sample size is small (e.g., Lange, Little & Taylor, 1989).

Briefly, we now suppose that the independently and identically distributed normal errors for model (2) are weighted by w_{ij}, so that observations with smaller weights are down-weighted. Given w_{ij} (and $\boldsymbol{\beta}_j$, σ^2), y_{ij} is distributed normal with variance (σ^2/w_{ij}). Additionally, w_{ij} is assumed to be distributed gamma, or $w_{ij} \sim \chi^2_\lambda/\lambda$. The results given above hold for a revised Gibbs sampler, and are augmented by the full conditional for the weights

$$w_{ij} \sim G\left\{(\lambda + 1)\big/2,\, 2\left[\left(y_{ij} - \mathbf{x}_{ij}^{k_j\prime}\boldsymbol{\beta}_j\right)^2\right]\big/\sigma^2 + \lambda^{-1}\right\}.$$

Because the expected value of the individual weight w_{ij} is inversely proportional to the square of a standardized residual, data more distant from the predicted values will count less for a specified degree of freedom. This weighting is amplified as we reduce λ. Seltzer, Novak, Choi, and Lim (in press) explored this strategy in an intervention study in order to accommodate some unusually low achieving students nested in remedial reading classrooms.[3]

Further Results

We now present results from weighting school-level regressions with (a) information about the precision of the school mean from its sample size, (b) ts with 4 and 11 degrees of freedom to evaluate the importance of outlying data points, and (c) their combination-case weighting of t-distributed observations at the school level. With

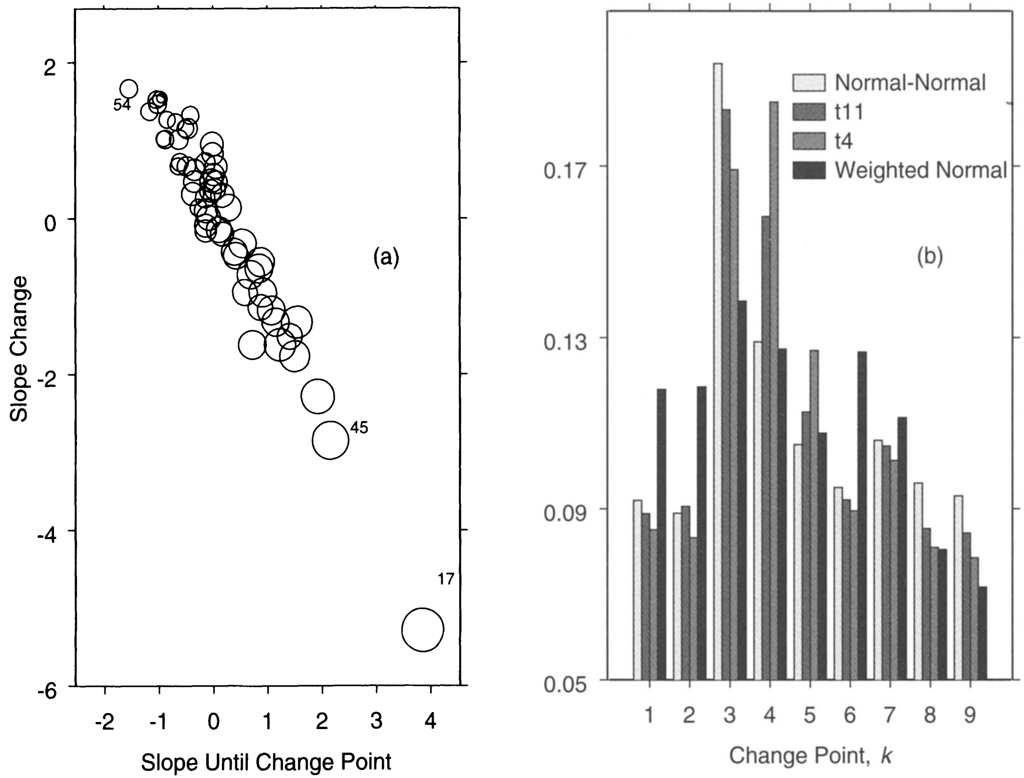

FIGURE 7. *(a) Scatterplot of posterior means under multilevel analysis: Change in slopes,* $(\hat{\beta}_{2j} - \hat{\beta}_{1j})$ *vs slope before change,* $\hat{\beta}_{1j}$*, with size of symbol proportional to join point,* $\hat{\beta}_{0j}$*. (b) Posterior distribution of estimated join points for some alternative multilevel join point models.*

reasonable adjustments to the hyper-priors previously employed in the unweighted analysis, all modifications to the Gibbs sampling procedure detailed above produced convergence after 30,000 updates.

Marginal posterior distributions of join points for school-level regressions employing *t*s with 4 and 11 degrees of freedom is shown in Figure 7b, and a normal-normal model employing case weights. Further details are omitted for brevity. Results suggest overall agreement between the normal-normal and the t_{11}-normal model, not unexpected because a t_{11} density approaches the normal, identifying $k = 3$ as the join point when change occurred for almost 19% of the schools in the system. A model using t_4 errors however suggests that closer to 20% of the schools changed but at $k = 4$, a year later. When analyzing our data with case weights using the normal-normal model, a join point for the system change is less distinctive.

We compare the relative fits to the data of the alternative models using naïve Bayes factor computations via Schwarz's criterion (Kass & Raftery, 1995).[4] For the unweighted data, the model with t_4 errors produced a better fit than either the normal-normal model or the model with t_{11} errors. For the weighted data, the normal-normal model fit the data better than the t_4-normal model, and even better than the t_{11}-normal model. This suggests that the relative instability of the within-school piecewise regression due to a small number of time points in the series can be mitigated with enough data for each time point.

Figure 8 shows the fits of various models for four selected schools. Plotted against the horizontal axis are the locations of the posterior mode of individual join point for (1) (unilevel) maximum likelihood, (2) Carlin et al. (1992) (unilevel) hierarchical Bayes, (3) normal-normal multilevel join point, (4) weighted normal-normal multilevel join point, (5) t_4-normal multilevel join point, and (6) t_{11}-normal multilevel join point solution. Overlaying the observed data are fitted curves from the normal-normal, the weighted normal-normal, and the t_4-normal models. Many solutions are typically consistent across models, the fits to data for School 11 above may be suspect if one were to focus on this school on its own. This is likely to be the result of excessive shrinkage, as suggested by a more reasonable fit from a separate hierarchical Bayes solution for each school. The effect of shrinkage is potentially a serious concern for interpretation. A more thorough analysis needs to identify all such schools for further investigation.

Conclusion and Discussion

In this article, we have illustrated an analysis for detecting changes in multiple time series for a multilevel setting. We have shown how it can be employed for evaluating intervention effects associated with a treatment to which the lengths of delay in the response may be unique to each treated units (see, e.g., Campbell & Stanley, 1963, pp. 37–43, for other possible outcome patterns). We argue that a more realistic description of change is possible when using an approach which neither assumes that the change point for a unit is known nor that units change synchronously. Thus our approach is in stark contrast to the presumption that the change point is known and is coincident with the time of intervention, the modus

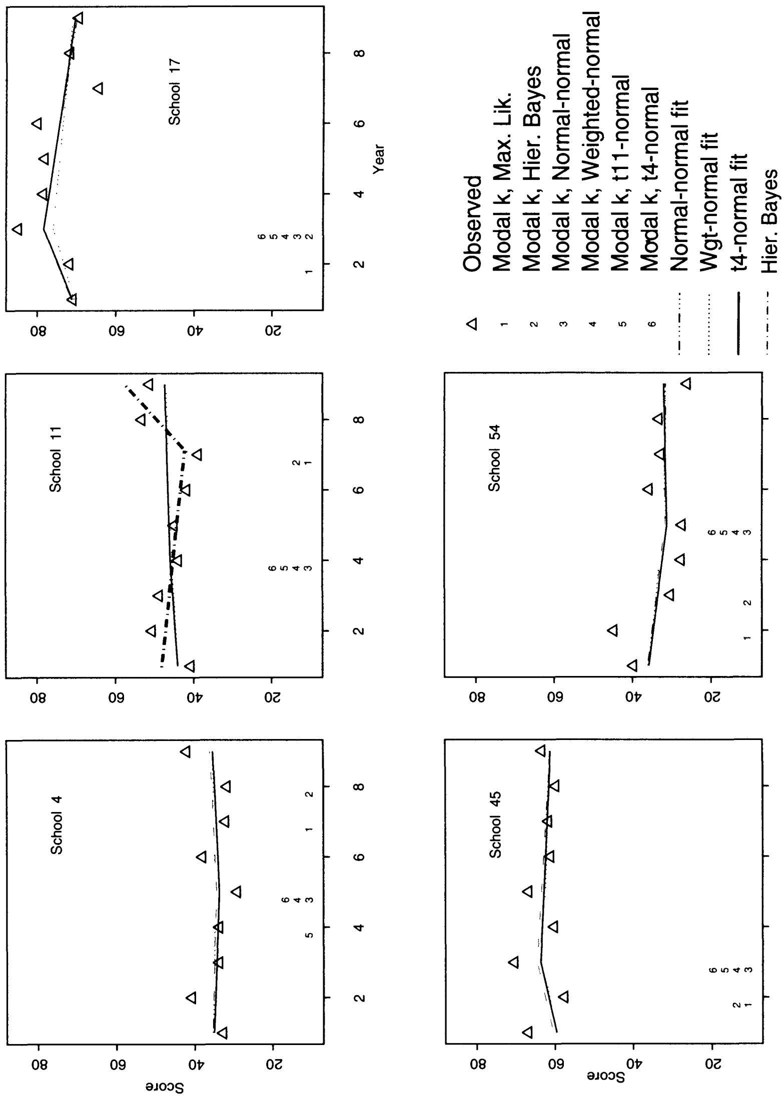

FIGURE 8. *Observed, fitted values with modal join point estimates based on alternative join-point models for selected schools.*

operandi in educational evaluation (e.g., research on the effects of class size reduction by Nye, Hedges, and Konstantopolous, 1999; Hanushek, 1999) and in behavioral research (in single subject designs, see Sidman, 1960; Barlow & Hersen, 1984; Crosbie, 1995), irrespective of the particular methodological strain employed for analysis. Additionally, we may also have an interest, in such situations, in estimating and understanding the time a treated unit takes to respond. For example, when the onset of a medical condition is not directly observed, detecting a change in a marker for the condition may provide a basis for inferring the time of onset, leading to a better description of the condition (see, e.g., Slate & Turnbull, 2000).

Other methods have been proposed in the past for determining the timing of important events. For example, if the timings of critical events (such as onset of drug use by minors in a particular urban community) are observed for all or most units, and we wish to estimate the time of onset for the collection of units (in order to improve the intervention), survival analysis may be helpful in determining average time of onset, recidivism, recovery relapse, and reoccurrence. Willett and Singer (1995) provided forceful arguments for considering such methods in educational modeling. However, survival analysis requires that the change event is itself observed for some of the units. If event occurrence is unobservable, as is the hallmark of our example above, both the timing (*when*) and the detectability (*whether*) of change must be inferred from the course of some observable marker of the unobserved process. In such situations, $\boldsymbol{\pi}'$, the posterior distribution of the join points, is particularly relevant.

We note briefly several avenues for future research in school effectiveness and accountability using the multilevel random join point model. To be an even more useful instrument for detecting and explaining change in educational processes, this model can easily be extended to accommodate the study of school *readiness* variables (covariates, e.g., teacher and principal turnover), in order to investigate their roles in the timing and the outcome of academic intervention. There remains, however, a static quality to the models treated here that is unsatisfactory. It should also be clear that our brief review is limited to the nonsequential change point problem and ignores, for reasons of scope and space, the significant research on monitoring sequential processes for changes (e.g., Smith, 1975). Finally, we also expect more work on detecting structural shifts in higher dimensional situations (Moen, Salazar, & Broemeling, 1985).

Notes

[1]The interested reader should also consult the original article for information on various adjustments made for student and school-grade demographic composition, as well as for a suspected test form effect.

[2]Like many similar studies of school performance currently underway, more and more information about students, their parents, teachers and schools are routinely added over time to this database to give a more complete portrayal of student development.

[3]We may also allow λ to vary by employing an adaptive *t* error distribution. Although this strategy renders our inferences independent of our choice of a particular λ value, it will produce wider interval estimates due to averaging over the uncertainly in λ.

[4]The reader is warned that this makes only for a rough comparison because the accuracy of Schwarz's criterion is unknown for heavy-tailed distributions.

References

Bacon, D. W., & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika, 58,* 525–534.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Ser. B, 37,* 149–192 (with discussion).

Bryk, A. S., Thum, Y. M., Easton, J. Q., & Luppescu, S. (1998). Assessing school academic productivity: The case of Chicago school reform. *Social Psychology of Education, 2,* 103–142.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Boston, MA: Houghton-Mifflin Co.

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics, 41,* 389–405.

Chib, S. (1998). Estimation and comparison of multiple change point models. *Journal of Econometrics, 86,* 221–241.

Cohen, A., & Kushary, D. (1994). Adaptive and unbiased predictors in a change point regression model. *Statistics and Probability Letters, 20,* 131–138.

Crosbie, J. (1995). Interrupted time-series analysis with short series: Why it is problematic; How it can be improved. In J. M. Gottman (Ed.), *The Analysis of Change* (pp. 361–395). Mahwah, NJ: Lawrence Erlbaum Associates.

DeGroot, M. H. (1970). *Optimal statistical decisions.* New York: McGraw-Hill.

El-Sayyad, G. M. (1975). A Bayesian analysis of the change point problem. *Egyptian Statistical Journal, 19,* 1–13.

Esterby, S. R., & el-Shaarawi, A. H. (1981). Inference about the point of change in a regression model. *Applied Statistics, 30,* 277–285.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85,* 398–409.

Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21,* 143–164.

Hinkley, D. V., Chapman, P., & Runger, G. (1980). *Change point problems* (Technical Report No. 382). University of Minneapolis, MN.

Hinkley, D. V., & Schechtman, E. (1987). Conditional bootstrap methods in the mean shift model. *Biometrika, 74,* 85–93.

Jandhyala, V. K., & Fotopoulos, S. B. (1999). Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika, 86,* 129–140.

Joseph, L., Wolfson, D. B., du Berger, R., & Lyle, R. M. (1996). Change point analysis of a randomized trial on the effects of calcium supplementation on blood pressure. In D. A.

Berry, & D. K. Stangl (Eds.), *Bayesian biostatistics* (pp. 617–649). New York: Marcel Dekker.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association, 84,* 881–896.

Maronna, R., & Yohai, V. J. (1978). A bivariate test for the detection of a systematic change in mean. *Journal of the American Statistical Association, 73,* 640–645.

Moen, D. H., Salazar, D., & Broemeling, L. D. (1985). Structural changes in multivariate regression models. *Communications in Statistics, A, 14,* 1757–1768.

Morrell, C. H., Pearson, J. D., Carter, H. B., & Brant, L. J. (1995). Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association, 90,* 45–53.

Müller, P., & Rosner, G. L. (1994). A semiparametric Bayesian model with hierarchical mixture priors. *Journal of the American Statistical Association, 92,* 1279–1292.

Nye, B., Hedges, L. V., & Konstantopolous, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21,* 127–142.

Picard, D. (1985). Testing and estimating change points in time series. *Advances in Applied Probability, 17,* 841–867.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association, 53,* 873–880.

Quandt, R. E., & Ramsey, J. N. (1978). Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association, 73,* 730–752 (with discussion).

Raftery, A. E., & Akman, V. E. (1986). Bayesian analysis of a Poisson process with a change point. *Biometrika, 73,* 85–89.

Seltzer, M. H., Novak, J., Choi, K. C., & Lim, N. (in press). Sensitivity analysis for hierarchical linear models employing *t* level-1 assumptions. *Journal of Educational and Behavioral Statistics.*

Sidman, M. (1960) *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Slate, E. H., & Clark, L. C. (1999). Using PSA to detect prostrate cancer onset: An application of Bayesian retrospective and prospective changepoint identification. In C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, & N. D. Singpurwalla (Eds.), *Case studies in Bayesian statistics IV* (pp. 511–534). New York: Springer-Verlag.

Slate, E. H., & Cronin, K. A. (1997). Changepoint modeling of longitudinal PSA as a biomarker for prostate cancer. In C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, & N. D. Singpurwalla (Eds.), *Case studies in Bayesian statistics III* (pp. 444–456). New York: Springer-Verlag.

Slate, E. H., & Turnbull, B. W. (2000). Statistical models for longitudinal biomarkers of disease onset. *Statistics in Medicine, 19,* 617–637.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change point in a sequence of random variables. *Biometrika, 62,* 407–416.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1995). *BUGS: Bayesian inference using Gibbs sampling,* Version 0.50. [Computer software.] Cambridge, UK: MRC Biostatistics Unit.

Stephens, D. A. (1994). Bayesian retrospective multiple changepoint identification. *Applied Statistics, 43,* 159–178.

Willett, J. B., & Singer, J. D. (1995). Investigating onset, cessation, relapse, and recovery: Using discrete-time survival analysis to examine the occurrence and timing of critical events. In J. M. Gottman (Ed.), *The analysis of change* (pp. 203–259). Mahwah, NJ: Lawrence Erlbaum Associates.

Wolfe, D. A., & Schechtman, E. (1984). Nonparametric statistical procedures for the change point problem. *Journal of Statistical Planning and Inference, 9,* 389–396.

Worsley, K. J. (1983). Testing for a two-phase multiple regression. *Technometrics, 25,* 35–42.

Zacks, S. (1983). Survey of classical and Bayesian approaches to the change point problem: Fixed sample and sequential procedures of testing and estimation. In M. Haseeb Rizvi, J. S. Rustagi, & D. Siegmund (Eds.), *Recent advances in statistics: Herman Chernoff festschrift* (pp. 245–269). New York: Academic Press.

Authors

YEOW MENG THUM is a Assistant Professor Department of Education, Graduate School of Education and Information Studies, 2019A Moore Hall, University of California, Los Angeles; thum@ucla.edu. He specializes in models for growth and change.

SUMAN K. BHATTACHARYA is a biostatistician at the Division of Clinical Biostatistics and Global Biometrics, R. W. Johnson Pharmaceutical Research Institute, 920 Route 202, PO Box 300, Raritan, NJ 08869; sbhatta4@prius.jnj.com. He specializes in Bayesian approaches to model building issues for repeated measurement data.

Received November 2000
Revision Received December 2000
Accepted October 2001

LINKED CITATIONS

You have printed the following article:

Detecting a Change in School Performance: A Bayesian Analysis for a Multilevel Join Point Problem

Yeow Meng Thum; Suman K. Bhattacharya

Journal of Educational and Behavioral Statistics, Vol. 26, No. 4. (Winter, 2001), pp. 443-468.

Stable URL:

http://links.jstor.org/sici?sici=1076-9986%28200124%2926%3A4%3C443%3ADACISP%3E2.0.CO%3B2-4

References

Estimating the Transition between Two Intersecting Straight Lines

David W. Bacon; Donald G. Watts

Biometrika, Vol. 58, No. 3. (Dec., 1971), pp. 525-534.

Stable URL:

http://links.jstor.org/sici?sici=0006-3444%28197112%2958%3A3%3C525%3AETTBTI%3E2.0.CO%3B2-J

Techniques for Testing the Constancy of Regression Relationships over Time

R. L. Brown; J. Durbin; J. M. Evans

Journal of the Royal Statistical Society. Series B (Methodological), Vol. 37, No. 2. (1975), pp. 149-192.

Stable URL:

http://links.jstor.org/sici?sici=0035-9246%281975%2937%3A2%3C149%3ATFTTCO%3E2.0.CO%3B2-7

Hierarchical Bayesian Analysis of Changepoint Problems

Bradley P. Carlin; Alan E. Gelfand; Adrian F. M. Smith

Applied Statistics, Vol. 41, No. 2. (1992), pp. 389-405.

Stable URL:

http://links.jstor.org/sici?sici=0035-9254%281992%2941%3A2%3C389%3AHBAOCP%3E2.0.CO%3B2-K

LINKED CITATIONS

Sampling-Based Approaches to Calculating Marginal Densities
Alan E. Gelfand; Adrian F. M. Smith
Journal of the American Statistical Association, Vol. 85, No. 410. (Jun., 1990), pp. 398-409.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199006%2985%3A410%3C398%3ASATCMD%3E2.0.CO%3B2-3

Some Findings from an Independent Investigation of the Tennessee STAR Experiment and from Other Investigations of Class Size Effects
Eric A. Hanushek
Educational Evaluation and Policy Analysis, Vol. 21, No. 2, Special Issue: Class Size: Issues and New Findings. (Summer, 1999), pp. 143-163.
Stable URL:
http://links.jstor.org/sici?sici=0162-3737%28199922%2921%3A2%3C143%3ASFFAII%3E2.0.CO%3B2-S

Conditional Bootstrap Methods in the Mean-Shift Model
David Hinkley; Edna Schechtman
Biometrika, Vol. 74, No. 1. (Mar., 1987), pp. 85-93.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198703%2974%3A1%3C85%3ACBMITM%3E2.0.CO%3B2-A

Capturing the Distributional Behaviour of the Maximum Likelihood Estimator of a Changepoint
Venkata K. Jandhyala; Stergios B. Fotopoulos
Biometrika, Vol. 86, No. 1. (Mar., 1999), pp. 129-140.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199903%2986%3A1%3C129%3ACTDBOT%3E2.0.CO%3B2-D

Bayes Factors
Robert E. Kass; Adrian E. Raftery
Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 773-795.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199506%2990%3A430%3C773%3ABF%3E2.0.CO%3B2-8

Robust Statistical Modeling Using the t Distribution
Kenneth L. Lange; Roderick J. A. Little; Jeremy M. G. Taylor
Journal of the American Statistical Association, Vol. 84, No. 408. (Dec., 1989), pp. 881-896.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198912%2984%3A408%3C881%3ARSMUTD%3E2.0.CO%3B2-U

LINKED CITATIONS

A Bivariate Test for the Detection of a Systematic Change in Mean
Ricardo Maronna; Victor J. Yohai
Journal of the American Statistical Association, Vol. 73, No. 363. (Sep., 1978), pp. 640-645.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197809%2973%3A363%3C640%3AABTFTD%3E2.0.CO%3B2-G

Estimating Unknown Transition Times Using a Piecewise Nonlinear Mixed-Effects Model in Men with Prostate Cancer
Christopher H. Morrell; Jay D. Pearson; H. Ballentine Carter; Larry J. Brant
Journal of the American Statistical Association, Vol. 90, No. 429. (Mar., 1995), pp. 45-53.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199503%2990%3A429%3C45%3AEUTTUA%3E2.0.CO%3B2-B

The Long-Term Effects of Small Classes: A Five-Year Follow-Up of the Tennessee Class Size Experiment
Barbara Nye; Larry V. Hedges; Spyros Konstantopoulos
Educational Evaluation and Policy Analysis, Vol. 21, No. 2, Special Issue: Class Size: Issues and New Findings. (Summer, 1999), pp. 127-142.
Stable URL:
http://links.jstor.org/sici?sici=0162-3737%28199922%2921%3A2%3C127%3ATLEOSC%3E2.0.CO%3B2-F

Testing and Estimating Change-Points in Time Series
Dominique Picard
Advances in Applied Probability, Vol. 17, No. 4. (Dec., 1985), pp. 841-867.
Stable URL:
http://links.jstor.org/sici?sici=0001-8678%28198512%2917%3A4%3C841%3ATAECIT%3E2.0.CO%3B2-U

The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes
Richard E. Quandt
Journal of the American Statistical Association, Vol. 53, No. 284. (Dec., 1958), pp. 873-880.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28195812%2953%3A284%3C873%3ATEOTPO%3E2.0.CO%3B2-4

LINKED CITATIONS

Estimating Mixtures of Normal Distributions and Switching Regressions
Richard E. Quandt; James B. Ramsey
Journal of the American Statistical Association, Vol. 73, No. 364. (Dec., 1978), pp. 730-738.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197812%2973%3A364%3C730%3AEMONDA%3E2.0.CO%3B2-6

Bayesian Analysis of a Poisson Process with a Change-Point
A. E. Raftery; V. E. Akman
Biometrika, Vol. 73, No. 1. (Apr., 1986), pp. 85-89.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198604%2973%3A1%3C85%3ABAOAPP%3E2.0.CO%3B2-Y

Sensitivity Analysis for Hierarchical Models Employing t Level-1 Assumptions
Michael Seltzer; John Novak; Kilchan Choi; Nelson Lim
Journal of Educational and Behavioral Statistics, Vol. 27, No. 2. (Summer, 2002), pp. 181-222.
Stable URL:
http://links.jstor.org/sici?sici=1076-9986%28200222%2927%3A2%3C181%3ASAFHME%3E2.0.CO%3B2-P

A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables
A. F. M. Smith
Biometrika, Vol. 62, No. 2. (Aug., 1975), pp. 407-416.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197508%2962%3A2%3C407%3AABATIA%3E2.0.CO%3B2-K

Bayesian Retrospective Multiple-Changepoint Identification
D. A. Stephens
Applied Statistics, Vol. 43, No. 1. (1994), pp. 159-178.
Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281994%2943%3A1%3C159%3ABRMI%3E2.0.CO%3B2-1

Testing for a Two-Phase Multiple Regression
K. J. Worsley
Technometrics, Vol. 25, No. 1. (Feb., 1983), pp. 35-42.
Stable URL:
http://links.jstor.org/sici?sici=0040-1706%28198302%2925%3A1%3C35%3ATFATMR%3E2.0.CO%3B2-X