A Dirichlet Process Functional Approach to Heteroscedastic-Consistent Covariance Estimation

George Karabatsos
University of Illinois-Chicago (UIC)

Thursday, June 16, 2016
Short session SH43 (2:30pm - 3:30pm):
Contributions in Bayesian and Approximate Bayesian Computation

Supported by NSF grant SES-1156372.
Introduction

• A line of research deals with the problem of determining the expression for the distribution of functionals of the Dirichlet process, with any prescribed error of approximation (Cifarelli & Regazzini, 1979; reviews: Regazzini et al., 2002; Lijoi & Prunster, 2009). Focused on mean/linear functionals of DP.

• I will show that the ordinary least squares (OLS) estimator, as a functional of the Mixture of Dirichlet process (MDP), has (approximately) posterior mean given by weighted least squares (WLS), and posterior covariance given by a weighted version of White’s (1980) heteroscedastic-consistent sandwich covariance estimator.

• This finding is based on a Taylor series approximation of a bootstrap distribution (i.e., Pólya urn) representation of the Dirichlet process (DP), using the multivariate delta method.

• Under a non-informative DP prior, this sandwich estimator closely approximates White's (1980) original (unweighted) sandwich covariance estimator.

• This work is a contribution to likelihood-free approximate Bayesian inference.
Introduction

• Moreover, for a particular choice of baseline distribution in the MDP, the WLS estimator corresponds to a generalized ridge regression estimator, which can handle:
  - multicollinear or singular covariate design matrices, including $p > n$ settings, while shrinking the coefficient estimates of irrelevant covariates towards zero;
  - Handle the fitting of nonlinear regressions via basis expansions.

• Bayesian ridge regression is tough to beat in terms of prediction (Griffin & Brown, 2013, *Bayesian Analysis*).

• The current study is the first to make connections between the DP and ridge regression, and the first to provide heteroscedastic-consistent posterior covariance estimation for the coefficients in ridge regression.

• I will then illustrate the MDP functional methodology through the analysis of simulated data sets, and two real data sets.

• The real data set illustrations will also include Vibration of Effects (VoE) analyses (Ioannidis, 2008: “Why most discovered true associations are inflated”, *Epidemiology*), to investigate how the covariate effect changes as a function of what other covariates are included in the model, and choice of prior.
Introduction

• All of these posterior quantities of this MDP functional methodology (the WLS and sandwich estimators) are analytically manageable and permit fast computations for large data sets.

• The DP prior was introduced with the motivation (Ferguson, 1973, p.209) that the DP posterior distribution:
  (1) "should be manageable analytically," and that
  (2) the "support of the prior distribution should be large-with respect to some suitable topology on the space of probability distributions on the sample space."

• This seminal paper then provided explicit analytical (closed form) solutions to a list of nonparametric statistical problems based on the DP posterior, including the estimation of a distribution function, median, quantiles, variance, covariance, and the probability that one variable exceeds another.

• The current study adds to this list, using the MDP functional methodology that will be presented in this talk.
MDP Model (prior)

- **Data:** \( Z_n = (z_i^T = (x_i^T, y_i))_{i=1}^n = (X, y) = (X_n, y_n) \)
  
  \( n \) observations of \( Z = (X, Y) \), sample space \( \mathcal{Z} = \{Z\} \subset \mathbb{R}^{K+1} \).

- **Clusters:** \( Z_{c_n}^* = (X_{c_n}^*, y_{c_n}^*) = (z_{c_n}^* = (x_{c_n}^T, y_{c_n}^*))_{c_n \leq n} \)

- **Cluster frequencies:** \( n_{c_n} = (n_1, \ldots, n_{c_n})^T \), with \( \sum_{c=1}^{c_n} n_c = n \).

- **Ridge baseline prior:**
  
  \( F_0(z) = N_{K+1}(x_i^T, y_i | m_z, V_z) \)
  
  MDP model:
  \( \alpha \sim \pi(\alpha), (\alpha > 0) \).

- **Unit ridge baseline prior:**
  \( v_{x_k} = 1, \text{ for } k = 2, \ldots, K \).

- **Ridge baseline prior:**
  
  \( F_0(z) = N(x_1 | 0, 0) \prod_{k=2}^{K} N(x_k | 0, v_{x_k})N(y | 0, 0) \)

  \( \pi(\alpha) \) is either a U(0, \( \xi \)) prior, or a truncated Cauchy-type prior:
  \( \pi(\alpha) = 1(0 < \alpha < \xi)/(\alpha + 1)^2, \alpha > 0 \) (Nandram & Yin, 2016).

- **Partitioning:**
  
  \( F|\alpha \sim \mathcal{D}\mathcal{P}(\alpha, F_0) \rightarrow F(B_1), \ldots, F(B_k) | \alpha \sim \text{Di}_k(\alpha F_0(B_1), \ldots, \alpha F_0(B_k)) \)
  
  all \( k \geq 1 \) partitions \( B_1, \ldots, B_k \) of \( \mathcal{Z} = \{Z\} \subset \mathbb{R}^{K+1} \).

- **Cluster allocation:**
  
  \( P(C_n = k | \alpha) = \frac{s_n(k)\alpha^k \Gamma(\alpha)}{\Gamma(\alpha + n)} \), \( (\alpha > 0) \)
  
  (the \( s_n(k) \) are the signless Stirling numbers.)
MDP Model (posterior)

- **Conditional posterior:**
  \[
  F(B_1), \ldots, F(B_k) \mid \mathbf{Z}_n, \alpha \sim \text{Di}_k(\alpha F_0(B_1) + \hat{n}F_n(B_1), \ldots, \alpha F_0(B_k) + \hat{n}F_n(B_k))
  \]
  all \( k \geq 1 \) partitions \( B_1, \ldots, B_k \) of \( \mathcal{Z} = \{\mathbf{Z}\} \subset \mathbb{R}^{K+1} \).

- **DP cond. posterior expectation (predictive; Pólya urn scheme):**
  \[
  \mathbb{E}[F(B) \mid \mathbf{Z}_n, \alpha] = \Pr(z_{n+1} \in B \mid \mathbf{Z}_n, \alpha)
  \]
  \[
  = \frac{n}{\alpha+n}\hat{F}_n(B) + \frac{\alpha}{\alpha+n}F_0(B) := \bar{F}_\alpha(B),
  \]
  \[
  = \sum_{c=1}^{c_n} \frac{n_c}{\alpha+n} \delta_{\mathbf{z}^*_c}(B) + \frac{\alpha}{\alpha+n}F_0(B)
  \]

- **DP cond. posterior variance:**
  \[
  \forall [F(B) \mid \mathbf{Z}_n, \alpha] = \frac{\bar{F}_\alpha(B)\{1 - \bar{F}_\alpha(B)\}}{\alpha + n + 1}, \forall B \in \mathcal{B}(\mathcal{Z})
  \]

- **\( \alpha \) posterior:**
  \[
  \pi(\alpha \mid c_n) \propto \pi(\alpha)\alpha^{c_n}\Gamma(\alpha)/\Gamma(\alpha + n)
  \]
Linear Regression Theory (Review)

- Data: \( Z_n = (z_i^T = (x_i^T, y_i))^n_{i=1} = (X, y) = (X_n, y_n) \)
- \( c_n < n \) dist. clusters: \( Z_{c_n}^* = (X_{c_n}^*, y_{c_n}^*) = (z_{c}^* = (x_{c}^*, y_{c}^*))_{c=1}^{c_n \leq n} \)
- Cluster frequencies: \( n_{c_n} = (n_1, \ldots, n_{c_n})^T \)

- Regression equation: \( Y_i = \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + \epsilon_i, \quad i = 1, \ldots, n. \)
- Regression errors: \( (\epsilon_1, \ldots, \epsilon_i, \ldots, \epsilon_n) \)
- Corresponding variances: \( \Phi = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \)
- Assuming exogeneity: \( \mathbb{E}[\epsilon_j | x_i] = 0, \quad i, j = 1, \ldots, n. \)

- OLS estimator: \( \hat{\beta} = (X^T X)^{-1} X^T y = (X_{c_n}^* \text{diag}(n_{c_n}) X_{c_n}^*)^{-1} X_{c_n}^* \text{diag}(n_{c_n}) y_{c_n}^* \)

- Sampling covariance matrix: \( \nabla(\hat{\beta}) = (X^T X)^{-1} X^T \Phi X (X^T X)^{-1} \)

- Under homoscedasticity: \( \nabla(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \) \( (\sigma^2 = \sigma_1^2 = \cdots = \sigma_n^2 \Phi = \sigma^2 I_n ) \)
\( \hat{\sigma}^2 (X^T X)^{-1}, \quad \hat{\sigma}^2 = (\frac{1}{n-K}) (y - X \hat{\beta})^T (y - X \hat{\beta}) \), consistent only under homoscedasticity.

- Sandwich estimator: \( \text{HC0} = (X^T X)^{-1} X^T \text{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2) X (X^T X)^{-1} \) (White, 1980)
\( \hat{u}_i = \hat{\epsilon}_i = y_i - x_i^T \hat{\beta}, \quad i = 1, \ldots, n \)
Consistent under hetero/homoscedasticity. \( \text{HC0} = \hat{\sigma}^2 (X^T X)^{-1} \) under homosc.

\( n^{1/2} (\hat{\beta}_n - \beta) \xrightarrow{d} N_K(0, n \nabla(\hat{\beta})) \), \( n \to \infty. \)
Matrix Identities

• **Observed data:** \( \mathbf{Z}_n = (\mathbf{z}_i = (x_i^T, y_i))^n_{i=1} = (\mathbf{X}, \mathbf{y}) = (\mathbf{X}_n, \mathbf{y}_n) \)
  has empirical dist. \( \hat{F}_n \) (and clusters \( \mathbf{Z}^{*\circ}_{c_n} = (\mathbf{X}^{\circ*}_{c_n}, \mathbf{y}^{\circ*}_{c_n}) = (\mathbf{z}^{\circ*}_c = (\mathbf{x}^{\circ*}_c, \mathbf{y}^{\circ*}_c))^{c_n \leq n}_{c=1} \)
  with mean \( \mathbf{m}_z = (\mathbf{m}_x^T, m_y)^T \) and \((K+1) \times (K+1)\) covariance matrix \( \hat{V}_z \),
  including the \( K \times K \) covariance matrix \( \hat{V}_x \) of \( \mathbf{X} \),
  and the \( K \times 1 \) vector \( \hat{V}_{x\mathbf{y}} \) of covariances between \( \mathbf{X} \) and \( \mathbf{y} \) (resp.).

• **Imaginary data:** \( \tilde{\mathbf{Z}}_S = (\mathbf{X}_S, \mathbf{y}_S) = ((\bar{x}_sk)_{s \times K}, (\bar{v}_{sk})_{s \times 1}) \) with eff. sample size \( \alpha \),
  has MDP baseline distribution, \( F_0 \), with mean \( \mathbf{m}_z = (\mathbf{m}_x^T, m_y)^T \),
  with \((K+1) \times (K+1)\) covariance matrix \( V_z \), with \( K \times K \) covariance matrix \( V_x \) of \( \mathbf{X} \),
  and the \( K \times 1 \) vector \( V_{x\mathbf{y}} \) of covariances between \( \mathbf{X} \) and \( \mathbf{y} \).

• The OLS estimator from the combined data set, \( (\bar{X}, \bar{Y}) = \left( \begin{array}{c} \mathbf{X}^{*\circ}_{c_n} \\ \mathbf{y}^{*\circ}_{c_n} \end{array} \right) \), is WLS:
  \[
  \hat{\beta} = \left[ \left( \begin{array}{c} \mathbf{X} \\ \mathbf{X}_S \end{array} \right)^T \text{diag}(1_n^T, (\alpha_S)I_S) \left( \begin{array}{c} \mathbf{X} \\ \mathbf{X}_S \end{array} \right) \right]^{-1} \left( \begin{array}{c} \mathbf{X} \\ \mathbf{X}_S \end{array} \right)^T \text{diag}(1_n^T, (\alpha_S)I_S) \left( \begin{array}{c} \mathbf{y} \\ \mathbf{y}_S \end{array} \right) \\
  = \left[ \left( \begin{array}{c} \mathbf{X}^{*\circ}_{c_n} \\ \mathbf{X}_S \end{array} \right)^T \text{diag}(n_{c_n}^T, (\alpha_S)I_S) \left( \begin{array}{c} \mathbf{X}^{*\circ}_{c_n} \\ \mathbf{X}_S \end{array} \right) \right]^{-1} \left( \begin{array}{c} \mathbf{X}^{*\circ}_{c_n} \\ \mathbf{X}_S \end{array} \right)^T \text{diag}(n_{c_n}^T, (\alpha_S)I_S) \left( \begin{array}{c} \mathbf{y}^{*\circ}_{c_n} \\ \mathbf{y}_S \end{array} \right) \\
  = \left( \mathbf{X}^T \text{diag}(n^T, (\alpha_S)I_S) \mathbf{X} \right)^{-1} \mathbf{X}^T \text{diag}(n^T, (\alpha_S)I_S) \mathbf{Y} \\
  = \left( \mathbf{X}^T \left[ \frac{1}{\alpha_n^T} \text{diag}(n^T, (\alpha_S)I_S) \right] \mathbf{X} \right)^{-1} \mathbf{X}^T \left[ \frac{1}{\alpha_n^T} \text{diag}(n^T, (\alpha_S)I_S) \right] \mathbf{Y} \\
  = \left( n(\hat{V}_x + \hat{m}_x \hat{m}_x^T) + \alpha(\mathbf{V}_x + \mathbf{m}_x \mathbf{m}_x^T) \right)^{-1} \left( n(\hat{V}_{x\mathbf{y}} + \hat{m}_y \hat{m}_x^T) + \alpha(\mathbf{V}_{x\mathbf{y}} + m_y \mathbf{m}_x) \right) 
  \]
Matrix Identities

- The OLS estimator from the combined data set, \((\bar{X}, \bar{Y}) = (\bar{x}_S^{\ast}, \bar{y}_S^{\ast})\), is WLS:
  \[
  \hat{\beta} = \left( \begin{pmatrix} \bar{X} \\ \bar{X}_S \end{pmatrix} \right)^T \text{diag}(1^n, (\frac{\alpha}{S})I_S) \left( \begin{pmatrix} \bar{X} \\ \bar{X}_S \end{pmatrix} \right) \left[ \begin{pmatrix} \bar{X} \\ \bar{X}_S \end{pmatrix} \right]^T \text{diag}(1^n, (\frac{\alpha}{S})I_S) \left( \begin{pmatrix} \bar{Y} \\ \bar{Y}_S \end{pmatrix} \right)
  \]
  \[
  = \left( \begin{pmatrix} \bar{X}^* \\ \bar{X}_S \end{pmatrix} \right)^T \text{diag}(n^n_{cn}, (\frac{\alpha}{S})I_S) \left( \begin{pmatrix} \bar{X}^* \\ \bar{X}_S \end{pmatrix} \right) \left[ \begin{pmatrix} \bar{X}^* \\ \bar{X}_S \end{pmatrix} \right]^T \text{diag}(n^n_{cn}, (\frac{\alpha}{S})I_S) \left( \begin{pmatrix} \bar{Y}^* \\ \bar{Y}_S \end{pmatrix} \right)
  \]
  \[
  = \left( \bar{X}^T \text{diag}(n^T, (\frac{\alpha}{S})I_S) \bar{X} \right)^{-1} \bar{X}^T \text{diag}(n^T, (\frac{\alpha}{S})I_S) \bar{Y}
  \]
  \[
  = \left( \bar{X}^T \left[ \begin{pmatrix} \frac{1}{\alpha+n} \text{diag}(n^T, (\frac{\alpha}{S})I_S) \end{pmatrix} \right] \bar{X} \right)^{-1} \bar{X}^T \left[ \begin{pmatrix} \frac{1}{\alpha+n} \text{diag}(n^T, (\frac{\alpha}{S})I_S) \end{pmatrix} \right] \bar{Y}
  \]
  \[
  = \left( n(\hat{V}_x + \hat{m}_x\hat{m}_x^T) + \alpha(V_x + m_xm_x^T) \right)^{-1} \left( n(\hat{V}_{xy} + \hat{m}_y\hat{m}_x) + \alpha(V_{xy} + m_ym_x) \right)
  \]
  \[
  \text{Diagonal elements of } (\alpha / S)I_S \text{ sum to } \alpha, \text{ the prior sample size of the DP.}
  \]

- If \(\alpha\) is a positive integer with \(S = \alpha\), then \((\alpha / S)I_S = (\alpha / \alpha)I_\alpha = I_\alpha\), and
  \[
  \hat{\beta} = (X_{n+\alpha}^T X_{n+\alpha})^{-1} X_{n+\alpha}^T y_{n+\alpha}, \text{ where } (X_{n+\alpha}, y_{n+\alpha}) = \left( \begin{pmatrix} X_n \\ \bar{X}_a \\ \bar{y}_a \end{pmatrix} \right).
  \]
Matrix Identities

- The OLS estimator from the combined data set, \((\bar{X}, \bar{Y}) = (\frac{X_{cn}}{\bar{x}_S}, \frac{Y_{cn}}{\bar{y}_S})\), is WLS:
  \[
  \hat{\beta} = \left[ \left( \frac{X}{\bar{x}_S} \right)^T \text{diag}(1_n, \frac{\alpha}{S} I_S) \left( \frac{X}{\bar{x}_S} \right) \right]^{-1} \left( \frac{X}{\bar{x}_S} \right)^T \text{diag}(1_n, \frac{\alpha}{S} I_S) \left( \frac{Y}{\bar{y}_S} \right)
  \]
  \[
  = \left[ \left( \frac{X_{cn}^*}{\bar{x}_S} \right)^T \text{diag}(n_{cn}^*, \frac{\alpha}{S} I_S) \left( \frac{X_{cn}^*}{\bar{x}_S} \right) \right]^{-1} \left( \frac{X_{cn}^*}{\bar{x}_S} \right)^T \text{diag}(n_{cn}^*, \frac{\alpha}{S} I_S) \left( \frac{Y_{cn}}{\bar{y}_S} \right)
  \]
  \[
  = \left( X^T \text{diag}(n^T, \frac{\alpha}{S} I_S) \bar{X} \right)^{-1} X^T \text{diag}(n^T, \frac{\alpha}{S} I_S) \bar{Y}
  \]
  \[
  = \left( X^T \left[ \frac{1}{\alpha+n} \text{diag}(n^T, \frac{\alpha}{S} I_S) \right] \bar{X} \right)^{-1} X^T \left[ \frac{1}{\alpha+n} \text{diag}(n^T, \frac{\alpha}{S} I_S) \right] \bar{Y}
  \]
  \[
  = \left( n(\hat{V}_x + \hat{m}_x \hat{m}_x^T) + \alpha(V_x + m_x m_x^T) \right)^{-1} \left( n(\hat{V}_{xy} + \hat{m}_y \hat{m}_x) + \alpha(V_{xy} + m_y m_x) \right)
  \]

- For any \(\alpha > 0\), possibly \(S \neq \alpha\), an extension of the fractional imputation method (e.g., Kim & Kim, 2012) can be used to simulate the imaginary data set.

- Draw a large number \((S)\) of samples \(z_{cn+1,s} = (x_{cn+1,s}, y_{cn+1,s}) \sim F_0\), for \(s = 1, \ldots, S\).

- \(\text{plims}_{S \to \infty} \frac{1}{S} \sum_{s=1}^{S} z_{cn+1,s} z_{cn+1,s}^T = (V_z + m_z m_z^T)\) by the law of large numbers.

- Then plug in \(\bar{X}_S := (x_{cn+1,s}^T)_{s=1}^{S}, \bar{y}_S := (y_{cn+1,s})_{s=1}^{S}\), into the OLS estimator above.
Under the ridge baseline prior, the OLS estimator also satisfies the equalities:

\[
\hat{\beta} = \left[ \left( \frac{X^*_{cn}}{X_K} \right)^T \text{diag}(n_{cn}^T, \alpha I_K) \left( \frac{X^*_{cn}}{X_K} \right) \right]^{-1} \left( \frac{X^*_{cn}}{X_K} \right)^T \text{diag}(n_{cn}^T, \alpha I_K) \left( \frac{y^*_{cn}}{y_K} \right)
\]

\[
= \left[ \left( \frac{n_{cn}^{1/2} x^*_{ck}}{\alpha^{1/2} \text{diag}(0, v_{x_2}, \ldots, v_K)^{1/2}} \right)^T \left( \frac{n_{cn}^{1/2} x^*_{ck}}{\alpha^{1/2} \text{diag}(0, v_{x_2}, \ldots, v_K)^{1/2}} \right) \right]^{-1} \left( \frac{n_{cn}^{1/2} y^*_{cn}}{0_K} \right)
\]

\[
= \left[ \left( \frac{X}{\alpha^{1/2} \text{diag}(0, v_{x_2}, \ldots, v_K)^{1/2}} \right)^T \left( \frac{X}{\alpha^{1/2} \text{diag}(0, v_{x_2}, \ldots, v_K)^{1/2}} \right) \right]^{-1} \left( \frac{y}{0_K} \right),
\]

\[X_K = \text{diag}(V_x + m_x m_x^T)^{1/2} = \text{diag}(0, v_{x_2}, \ldots, v_{x_K})^{1/2};\]

\[y_K = (V_{xy} + m_y m_x)^{1/2} = (0_K + 0_K)^{1/2} = 0_K\]

- Uses imaginary data \((X_K, y_K)\), obtained by deterministic single imputation.
- The OLS solution above, obtained by the MDP ridge baseline prior, is the \textbf{generalized ridge regression estimator} with shrinkage parameters \(\alpha \cdot (0, v_2, \ldots, v_K)\); and if further \(v_2 = \cdots = v_K = 1\), then the OLS solution above is the \textbf{ridge regression estimator} (Hoerl & Kennard, 1970) with shrinkage parameter \(\alpha\) (Hastie et al., 2009, p. 96, observed the latter without any reference to the DP).
- Under a ridge baseline prior, we may just use: \((\bar{X}, \bar{Y}) = \left( \frac{X^*_{cn}}{\bar{X}_K}, \frac{y^*_{cn}}{\bar{y}_K} \right)\). 11
**MDP Bootstrap Procedure**

- Given $\alpha$, a bootstrap draw $b$ of a random c.d.f. $F^*$ can be constructed by taking:

$$
F_b^*(t) = \sum_{c=1}^{c_n+1} \frac{n_{cb}^*}{n+\alpha+1} 1(z_c^* \leq t) + \frac{(\alpha-\text{floor}(\alpha))}{n+\alpha+1} 1(z_{c_n+1}^* (\text{ceil}(\alpha)>\alpha), b \leq t)
$$

given $n + \text{ceil}(\alpha) + 1$ draws $z_{cb}^*$ from $\mathbb{E}[F(\cdot) | Z_n, \alpha] = \sum_{c=1}^{c_n} \frac{n_c}{\alpha+n} \delta_{z_c^*}(\cdot) + \frac{\alpha}{\alpha+n} F_0(\cdot)$ and a posterior draw $\alpha \sim \pi(\alpha | Z_n)$.

- This bootstrap procedure is the same as taking a multinomial draw:

$$
n_{c, n+1, b}^* = (n_{cb}^*)_{(c_n+1) \times 1} \sim \text{Mult}(n + \text{ceil}(\alpha) + 1; \frac{n_1}{\alpha+n}, \ldots, \frac{n_c}{\alpha+n}, \ldots, \frac{n_{cn}}{\alpha+n}, \frac{\alpha}{\alpha+n})
$$

with a draw $z_c^*$ made with probability $n_c / (\alpha + n)$, for $c = 1, \ldots, c_n$, and a draw $z_{c(n)+1}^* \sim F_0$ made with probability $\alpha / (\alpha + n)$.

- This bootstrap sampling process may be repeated for $b = 1, \ldots, B$, for large $B$.

- Then:
  - $\mathbb{E}[F^*(B) | Z_n, \alpha] \doteq \mathbb{E}[F(B) | Z_n, \alpha]$, $\forall B \in \mathcal{B}(\mathbb{Z})$ ($\doteq$ denotes near equality);
  - $\mathbb{V}[F^*(B) | Z_n, \alpha] \doteq \mathbb{V}[F(B) | Z_n, \alpha]$, $\forall B \in \mathcal{B}(\mathbb{Z})$;
  - $F$ has twice the skewness of $F^*$, but is small anyways;
  - $G(\varphi(F) | Z_n, \alpha) \doteq G^*(\varphi(F^*) | Z_n, \alpha)$, (c.d.f.s), any well-behaved functional $\varphi$;
  - All above are true marginally over the posterior $\pi(\alpha | Z_n)$.
MDP Bootstrap Procedure

• For the OLS functional, we may adopt the bootstrap sampling scheme:

\[
\hat{\beta}(F^*) = \hat{\beta}(n^*) = (X^\top \text{diag}(n^*)X)^{-1}X^\top \text{diag}(n^*)Y,
\]

\[
n^* = \frac{n+\alpha+1}{n+\text{ceil}(\alpha)+1}(n_{\text{cb}})_{(c_n+1)\times 1}
\]

\[
(n_{\text{cb}})_{(c_n+1)\times 1} | Z_n, \alpha \sim \text{Mu}_{c_n+S}(n + \text{ceil}(\alpha) + 1; \frac{n_1}{\alpha+n}, \ldots, \frac{n_{cn}}{\alpha+n}, \frac{\alpha/S}{\alpha+n} \otimes 1_S^\top),
\]

\[
\alpha | Z_n \sim \pi(\alpha | c_n).
\]

• The bottom \( S \) (or \( K \)) rows of \( (X, Y) \) already contain the imaginary observations sampled from the MDP baseline, \( F_0 \).

• The factor \((n + \alpha + 1) / (n + \text{ceil}(\alpha) + 1)\) addresses continuous \( \alpha \).

• We have:

\[
\mathbb{E}(n^* | Z_n, \alpha) = \bar{n}_\alpha^* = (n + \alpha + 1)(\frac{n_1}{\alpha+n}, \ldots, \frac{n_{cn}}{\alpha+n}, \frac{\alpha/S}{\alpha+n} \otimes 1_S^\top)^\top,
\]

\[
\nabla(n^* | Z_n, \alpha) = \text{diag}(\bar{n}_\alpha^*) - (n + \alpha + 1)^{-1}\bar{n}_\alpha^*\bar{n}_\alpha^\top.
\]

• For a ridge baseline prior, use \((\alpha/(\alpha + n)) \otimes 1_K^\top\) instead of \(\{\alpha/S\}/(\alpha + n) \otimes 1_S^\top\).

• Lancaster (2003) made rather similar observations as above (without using \( \alpha \)):

Classical (Efron) bootstrap uses: \( n^* \sim \text{Multinomial}(n; n_1/n, \ldots, n_{c(n)}/n) \);
Bayesian bootstrap uses: \( n^* \sim \text{Dirichlet}(1_1, \ldots, 1_n) \).
• Recall: \( \mathbb{E}(n^* | Z_n, \alpha) = \mathbf{n}_{\alpha}^* = (n + \alpha + 1)(\frac{n_1}{\alpha + n}, \ldots, \frac{n_{cn}}{\alpha + n}, \frac{\alpha/S}{\alpha + n} \otimes \mathbf{1}_S^T) \),

\( \nabla(n^* | Z_n, \alpha) = \text{diag}(\mathbf{n}_{\alpha}^*) - (n + \alpha + 1)^{-1}\mathbf{n}_{\alpha}^* \mathbf{n}_{\alpha}^{*\top} \).

• Again, for a ridge baseline prior, use \( (\alpha/(\alpha + n)) \otimes \mathbf{1}_K^\top \) instead of \( (\{\alpha/S\}/(\alpha + n)) \otimes \mathbf{1}_S^\top \).

• \( \pi(\alpha) \) is either a \( U(0, \xi) \) prior, or a Cauchy type prior truncated above by \( \xi \). Let \( A \) be a fine grid of \( \alpha \) defined over the support of the prior, \( \pi(\alpha) \). (The grid \( A \) may range from \( .005 \) to \( \xi \) in intervals of \( .005 \)).

• Then, marginalizing over the posterior, \( \pi(\alpha | c_n) \), and by the total law of probability for expectations and covariances, the marginal expectation and covariance matrix can be approximated and rapidly computed by:

\[
\begin{align*}
\mathbb{E}(n^* | Z_n) &= \mathbf{n}^* \approx \sum_{\alpha \in A} \mathbb{E}(n^* | Z_n, \alpha) \pi(\alpha | c_n), \\
\nabla(n^* | Z_n) &\approx \sum_{\alpha \in A} \nabla(n^* | Z_n, \alpha) \pi(\alpha | c_n) + \sum_{\alpha \in A} \mathbb{E}(n^* | Z_n, \alpha) \{\mathbb{E}(n^* | Z_n, \alpha)\}^\top \pi(\alpha | c_n) \\
&\quad - \mathbb{E}(n^* | Z_n) \{\mathbb{E}(n^* | Z_n)\}^\top.
\end{align*}
\]
Multivariate Delta Method

- We now consider a deterministic approach to evaluating the distribution of a functional (e.g., $\beta^*(F^*)$) of the MDP posterior, using a prescribed error of approximation (as in research on DP functionals; Regazzini, et al. 2002).
- We employ the multivariate delta method to approximate the (MDP bootstrap) posterior distribution of $\beta(F^*)$ via a Taylor series approximation $\beta^*(n^*) \approx \beta(n^*)$ of $\beta(n^*)$ around the posterior mean, $\bar{n}^*$. With $\partial \beta(n) / \partial n$ a $K \times (c_n + S)$ matrix of first derivatives ($S = K$ for ridge), this Taylor series approximation is given by:

$$
\beta^*(n^*) = \beta(\bar{n}^*) + \left[ \frac{\partial \beta(n^*)}{\partial n^*} \right]_{n=\bar{n}^*} (n^* - \bar{n}^*)
$$

$$
= \beta(\bar{n}^*) + \left[ (Y - X\beta(\bar{n}^*))^T \otimes (X^T \text{diag}(\bar{n}^*)X)^{-1}X^T \right] \frac{\partial \text{vec}\{\text{diag}(n)\}}{\partial (n)^T} (n - \bar{n}^*)
$$

$$
= \beta(\bar{n}^*) + \left[ u(\bar{n}^*)^T \otimes \{X(X^T \text{diag}(\bar{n}^*)X)^{-1}\}^T \right] (n - \bar{n}^*)
$$

$$
= \beta(\bar{n}^*) + R(\bar{n}^*)^T (n^* - \bar{n}^*),
$$

where: $u(\bar{n}^*) = Y - X\beta(\bar{n}^*)$; $e_c = (1(c = 1), \ldots, 1(c = c_n + S))^T$; $R(\bar{n}^*) = (e_1 e_1^T, \ldots, e_{c_n+S} e_{c_n+S}^T)(u(\bar{n}^*) \otimes X(X^T \text{diag}(\bar{n}^*)X)^{-1})$

$$
= \text{diag}\{u(\bar{n}^*)\} X(X^T \text{diag}(\bar{n}^*)X)^{-1}, \text{ is } (c_n + S) \times K.
$$
Then the posterior distribution of $n^*$ implies that the approximation has exact posterior mean given by the WLS estimator:

$$E(\beta^*(n^*)|Z_n) = \beta(\bar{n}^*) = (X^T \text{diag}(\bar{n}^*)X)^{-1}X^T \text{diag}(\bar{n}^*)Y$$

and exact covariance matrix given by the heteroscedastic-consistent sandwich estimator for WLS (Greene, 2012, p. 319):

$$V(\beta^*(n^*)|Z_n) = \left[ \frac{\partial \beta(n^*)}{\partial n^*} \right]_{n=\bar{n}^*} V(n^*|Z_n) \left[ \frac{\partial \beta(n^*)}{\partial n^*} \right]_{n=\bar{n}^*}^T = R(\bar{n}^*)^T V(n^*|Z_n) R(\bar{n}^*)$$

$$= (X^T \text{diag}(\bar{n}^*)X)^{-1} [X^T \text{diag}(\bar{n}^* \circ \{u(\bar{n}^*)\}^2)X] (X^T \text{diag}(\bar{n}^*)X)^{-1}$$

where \circ denotes the Hadamard product operator. Poirier (2011, *Economic Reviews*) obtained a similar result.

Then the posterior variances from this matrix,

$$\text{diag}\{V(\beta^*(n)|Z_n)\} = \{V(\beta_1^*(n^*)|Z_n), \ldots, V(\beta_K^*(n^*)|Z_n)\},$$

provide the asymptotic-consistent 95% posterior credible interval,

$$\beta_k^*(n^*) \pm 1.96 \{V(\beta_k^*(n^*)|Z_n)\}^{1/2}, \text{ for } k = 1, \ldots, K.$$
Non-informative DP prior case

• Furthermore, suppose that:
  - the MDP model assumes a non-informative DP prior, defined by the specification $\alpha \to 0$;
  - The number of distinct cluster in the data is $c_n = n$, so that $\bar{n} = \bar{n}_0 = (1_n^T, 0_S^T)^T$ ($S = K$ for ridge baseline prior).

• Then the conditional posterior distribution of the MDP is Dirichlet (Di),
  $\theta_{c_n} = (\theta_1, \ldots, \theta_{c_n})^T | Z_n \sim \text{Di}_{c_n}(\mathbf{n}_{c_n})$, with support points the $c_n$ observed cluster values $\{z_c^\ast\}_{c=1}^{c_n \leq n}$, where the random probabilities, $\theta_c = \Pr(z = z_c^\ast)$, $c = 1, \ldots, c_n$, coincide with those of the Bayesian Bootstrap (Rubin, 1981).

• Then,
  - the DP posterior predictive distribution function reduces to:
    $\Pr(z_{n+1} \in B | Z_n) = \hat{F}_n(B) = \sum_{c=1}^{c_n} \frac{n_c}{n} \delta_{z_c^\ast}(B), \quad \forall B \in \mathcal{B}(\mathcal{Z})$, the distribution function used by the classical bootstrap (Efron, 1979);
  - the posterior mean (WLS) estimate of $\beta$ is approximately equal to the usual OLS estimator, $\hat{\beta}(\bar{n}_0) \approx (X^TX)^{-1}X^Ty$;
  - and the posterior covariance matrix of $\beta$ reduces to White’s (1980) heteroscedastic consistent covariance estimator:
    $\nabla(\beta^*(\mathbf{n}^*) | Z_n) = (X^TX)^{-1}[X^T \text{diag}(\{u(\bar{n}_0^*)\}^{-2})X](X^TX)^{-1} \approx HC0$. 

17
### Simulation Study (coverage rates)

<table>
<thead>
<tr>
<th>X dist.</th>
<th>n</th>
<th>$a_h = 0 (0)$</th>
<th>$a_h = 1 (.05)$</th>
<th>$a_h = 2 (.1)$</th>
<th>$a_h = 2.5 (.15)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>c u h</td>
<td>c u h</td>
<td>c u h</td>
<td>c u h</td>
</tr>
<tr>
<td>U(0,1)</td>
<td>10</td>
<td>.59 .58 .83</td>
<td>.72 .71 .82</td>
<td>.87 .87 .79</td>
<td>.93 .92 .78</td>
</tr>
<tr>
<td>Ex(1)</td>
<td>10</td>
<td>.78 .78 .79</td>
<td>.74 .74 .73</td>
<td>.70 .69 .68</td>
<td>.66 .66 .64</td>
</tr>
<tr>
<td>N(0,25)</td>
<td>10</td>
<td>.81 .81 .82</td>
<td>.65 .65 .67</td>
<td>.64 .64 .67</td>
<td>.66 .66 .70</td>
</tr>
<tr>
<td>AR(1)</td>
<td>10</td>
<td>.82 .82 .81</td>
<td>.81 .80 .80</td>
<td>.78 .79 .78</td>
<td>.77 .77 .77</td>
</tr>
<tr>
<td>U(0,1)</td>
<td>20</td>
<td>.74 .74 .90</td>
<td>.84 .84 .89</td>
<td>.92 .92 .87</td>
<td>.95 .95 .86</td>
</tr>
<tr>
<td>Ex(1)</td>
<td>20</td>
<td>.86 .86 .86</td>
<td>.81 .81 .80</td>
<td>.75 .75 .75</td>
<td>.71 .71 .70</td>
</tr>
<tr>
<td>N(0,25)</td>
<td>20</td>
<td>.88 .88 .88</td>
<td>.84 .84 .85</td>
<td>.88 .88 .89</td>
<td>.90 .90 .91</td>
</tr>
<tr>
<td>AR(1)</td>
<td>20</td>
<td>.88 .88 .89</td>
<td>.87 .87 .87</td>
<td>.86 .86 .86</td>
<td>.85 .85 .86</td>
</tr>
<tr>
<td>U(0,1)</td>
<td>50</td>
<td>.88 .88 .93</td>
<td>.92 .92 .93</td>
<td>.94 .94 .92</td>
<td>.94 .94 .92</td>
</tr>
<tr>
<td>Ex(1)</td>
<td>50</td>
<td>.90 .90 .91</td>
<td>.86 .86 .86</td>
<td>.82 .82 .82</td>
<td>.82 .82 .82</td>
</tr>
<tr>
<td>N(0,25)</td>
<td>50</td>
<td>.92 .92 .92</td>
<td>.93 .93 .93</td>
<td>.96 .96 .97</td>
<td>.98 .98 .98</td>
</tr>
<tr>
<td>AR(1)</td>
<td>50</td>
<td>.93 .93 .93</td>
<td>.92 .92 .92</td>
<td>.91 .91 .92</td>
<td>.91 .91 .92</td>
</tr>
<tr>
<td>U(0,1)</td>
<td>100</td>
<td>.92 .92 .94</td>
<td>.94 .94 .94</td>
<td>.94 .94 .94</td>
<td>.94 .94 .94</td>
</tr>
<tr>
<td>Ex(1)</td>
<td>100</td>
<td>.92 .92 .92</td>
<td>.89 .89 .90</td>
<td>.89 .89 .89</td>
<td>.90 .90 .90</td>
</tr>
<tr>
<td>N(0,25)</td>
<td>100</td>
<td>.94 .94 .94</td>
<td>.96 .96 .96</td>
<td>.98 .98 .98</td>
<td>.99 .99 .99</td>
</tr>
<tr>
<td>AR(1)</td>
<td>100</td>
<td>.94 .94 .94</td>
<td>.93 .93 .94</td>
<td>.93 .93 .93</td>
<td>.93 .93 .93</td>
</tr>
<tr>
<td>U(0,1)</td>
<td>500</td>
<td>.94 .94 .95</td>
<td>.95 .95 .95</td>
<td>.95 .95 .95</td>
<td>.94 .94 .95</td>
</tr>
<tr>
<td>Ex(1)</td>
<td>500</td>
<td>.95 .95 .95</td>
<td>.93 .93 .93</td>
<td>.95 .95 .95</td>
<td>.96 .96 .96</td>
</tr>
<tr>
<td>N(0,25)</td>
<td>500</td>
<td>.95 .95 .95</td>
<td>.98 .98 .98</td>
<td>1.0 1.0 1.0</td>
<td>1.0 1.0 1.0</td>
</tr>
<tr>
<td>AR(1)</td>
<td>500</td>
<td>.95 .95 .95</td>
<td>.95 .95 .95</td>
<td>.95 .95 .95</td>
<td>.95 .95 .95</td>
</tr>
</tbody>
</table>

**4 × 4 × 5 simulation design:**
- X distribution type (4), by heterosced. level (4), by sample size level (5).
- 10,000 data sets simulated for each of 80 conditions.

**Data Generating model:**
\[ y_i \mid x_i, \beta, \sigma_i^2 \sim N(\beta_0 + \beta_1 x_i, \sigma_i^2); \]
\[ \beta_0 = \beta_1 = 1, \]
\[ \sigma_i^2 = \exp(a_h x_i + a_h x_i^2), i = 1, \ldots, n. \]

**Models fitted to data:**
- c: MDP with ridge base and truncated Cauchy-type prior
  \[ \pi(\alpha) = 1(0 < \alpha < \xi)/(\alpha + 1)^2. \]
- u: MDP with ridge base and \( U(\alpha \mid 0, 3) \) prior.
- h: HC0 (White, 1980).

**Results:** Rates similar across models and \( \approx .95, \) esp. for larger \( n. \)
### Simulation Study (rate means, s.d.s, by condition)

<table>
<thead>
<tr>
<th></th>
<th>Heteroscedasticity Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_h = 0$ (0)</td>
</tr>
<tr>
<td></td>
<td>c</td>
</tr>
<tr>
<td><strong>U(0,1)</strong></td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
</tr>
<tr>
<td><strong>Ex(1)</strong></td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
</tr>
<tr>
<td><strong>N(0,25)</strong></td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
</tr>
<tr>
<td><strong>AR(1)</strong></td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
</tr>
<tr>
<td><strong>n = 10</strong></td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
</tr>
<tr>
<td><strong>n = 20</strong></td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
</tr>
<tr>
<td><strong>n = 50</strong></td>
<td>.91</td>
</tr>
<tr>
<td></td>
<td>(.02)</td>
</tr>
<tr>
<td><strong>n = 100</strong></td>
<td>.93</td>
</tr>
<tr>
<td></td>
<td>(.01)</td>
</tr>
<tr>
<td><strong>n = 500</strong></td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>(.00)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
</tr>
</tbody>
</table>

### Data Generating model:

$$y_i | x_i, \beta, \sigma_i^2 \sim N(\beta_0 + \beta_1 x_i, \sigma_i^2);$$

$$\beta_0 = \beta_1 = 1,$$

$$\sigma_i^2 = \exp(a_{ih} x_i + a_{ih} x_i^2), i = 1, \ldots, n.$$  

U(0,1) X dist: $a_h = 0, 1, 2, 2.5.$ Else: $a_h = 0, .05, .1, .15.$

### Models fitted to data:

**c:** MDP with ridge base and truncated Cauchy-type prior

$$\pi(\alpha) = 1(0 < \alpha < \zeta)/(\alpha + 1)^2.$$  

**u:** MDP with ridge base and U(\(\alpha \mid 0, 3\)) prior.

**h:** HC0 (White, 1980).

### Results:

Rates similar across models and $\approx .95$, esp. for larger $n$.  

### 4 x 4 x 5 simulation design:

- X distribution type (4), by heterosced. level (4), by sample size level (5).
- 10,000 data sets simulated for each of 80 conditions.
LMT Data Analysis Example

• $n = 347$ observations from undergraduate teacher education students (89.9% female) who each attended one of four Chicago universities between the Fall 2007 semester and Fall 2013 spring semesters, inclusive, excluding summers.
• Dependent variable ($Y$): math teaching ability score (25-item test of Learning Math for Teaching; LMT, 2012).
• Covariates: $Year$, $Year^2$, and $CTPP = 1(Year \geq 2010.9)$, an indicator of the administration of the new (versus old) teaching curriculum.
• All covariates rescaled to have mean zero and variance 1 before data analysis.

<table>
<thead>
<tr>
<th>$\beta(\bar{n}^*)$</th>
<th>pSD</th>
<th>ES</th>
<th>95%PI $\beta(\bar{n}^*)$</th>
<th>OLS $\hat{\beta}$</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.90</td>
<td>.18</td>
<td>72.76</td>
<td>(12.55, 13.24)</td>
<td>12.90</td>
</tr>
<tr>
<td>Year</td>
<td>−.69</td>
<td>.57</td>
<td>−1.21</td>
<td>(−1.81, .43)</td>
<td>−478.17</td>
</tr>
<tr>
<td>$Year^2$</td>
<td>−.65</td>
<td>.57</td>
<td>−1.14</td>
<td>(−1.78, .47)</td>
<td>476.75</td>
</tr>
<tr>
<td>CTPP</td>
<td>.60</td>
<td>.28</td>
<td>2.10</td>
<td>(.04, 1.15)</td>
<td>.67</td>
</tr>
</tbody>
</table>
• Vibration of effects (VoE) analysis (Ioannidis, 2008) provides a way to assess how much a covariate’s effect size differs (or vibrates) over different ways that the analysis can be done, e.g. with respect to different:
  - variables included and/or excluded in the analysis (statistical adjustments);
  - models used;
  - definitions of outcomes and predictors;
  - and inclusion and exclusion criteria for the study population.

• We will consider a VoE analysis to investigate how the effect size of the CTPP treatment variable ($T$),

$$\hat{ES}_{Ta} = \tilde{\beta}^{*}_T \sqrt{\nabla_{Ta}} = \tilde{\beta}^{*}_T(n^*_a) / \sqrt{\nabla(\beta^*_T(n^*_a) \mid Z_n, \alpha)}^{1/2}$$

varies as a function of the other covariates that are included in the regression model, and $\alpha$.

• $\tilde{\beta}$ is a WLS estimate that excludes a missing variable, $U$, from the regression. See next slide for more details (on bias analysis).

• Covariate selection is undertaken using the fast LARS algorithm (Efron et al. 2004) applied to the data augmented by the baseline prior imaginary observations.
LMT Data Analysis Example

• We may add a quantitative bias analysis into the VoE analysis.

• Let \((X, T)\) as covariates in the regression equation, incl. treatment variable \(T\), and let \(U\) be a missing covariate from the regression equation, a confounder such that \(\varepsilon_{iU} = \gamma u_i + \varepsilon_i (i = 1, \ldots, n)\) is correlated with \(T\) (i.e., hidden bias).

• If there is hidden bias, then least-squares estimates are inconsistent.

• If \((T, U, X)\) have no interactions, then we can write:

\[
\beta_T(\mathbf{n}_a^*) = \tilde{\beta}_T(\mathbf{n}_a^*) - \gamma \int \int \{udF_U(u \mid T = 1, x) - udF_U(u \mid T = 0, x)\} dF_X(x)
\]

(see VanderWeele & Arah, 2011)

where:

\(\tilde{\beta}_T(\mathbf{n}_a^*)\) is \(T\)'s WLS slope estimator in a regression fit that excludes \(U\) (given \(\alpha\)),

\(\beta_T(\mathbf{n}_a^*)\) is the WLS slope estimator in a regression fit that includes \(U\) (given \(\alpha\)).

• If the missing variable \(U\) is binary, we may evaluate how sensitive the effect size of \(T\) is, using the formula displayed above, where \(F_U( u \mid T = 1, x; \lambda)\) and \(F_U( u \mid T = 0, x; \lambda)\) are based on a binary logistic regression, over independent standard normal samples of \((\gamma, \lambda)\).
LMT Data Analysis Example

Vibration of Effects Analysis

Vibration of Effects on Algebra

605 regressions

\[ GIC_2 = \frac{y - x^2}{\sigma^2} \]

\[ \text{Z:CTPP Effect Size} \]

\[ \text{DP precision } \alpha \]

An Effect Size, \( ES = \beta / \sqrt{V}_{HC} \), heteroscedastic-consistent (HC).

A random biased \( ES \).

Red: Median \( GIC \) (and 95\% CI) for a \( ES \) significance level.

\# covariates in regression
PIRLS Data Analysis Example

- Data on \( n = 565 \) low-income students from 21 U.S. elem schools (PIRLS 2006).
- Dependent variable: Student literacy score (zREAD).
- 8 covariates: student male status (1 if MALE, or 0); AGE; class size (SIZE); class percent of English language learners (ELL); teacher years of experience (TEXP4) and education level (EDLEVEL = 5 if bachelor's; = 6 if at least master's); school enrollment (ENROL), safety (SAFE=1 high; SAFE = 3 low).
- Each variable in the data set was rescaled to have mean 0 and variance 1.

<table>
<thead>
<tr>
<th>( \beta(n^*) )</th>
<th>pSD</th>
<th>ES</th>
<th>95%PI ( \beta(n^*) )</th>
<th>OLS ( \hat{\beta} )</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−.48</td>
<td>.04</td>
<td>−12.67</td>
<td>(−.56, −.41)</td>
<td></td>
</tr>
<tr>
<td>MALE</td>
<td>−.06</td>
<td>.04</td>
<td>−1.44</td>
<td>(−.13, .02)</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>−.20</td>
<td>.04</td>
<td>−5.22</td>
<td>(−.27, −.12)</td>
<td></td>
</tr>
<tr>
<td>SIZE</td>
<td>−.07</td>
<td>.05</td>
<td>−1.48</td>
<td>(−.16, .02)</td>
<td></td>
</tr>
<tr>
<td>ELL</td>
<td>−.13</td>
<td>.05</td>
<td>−2.81</td>
<td>(−.22, −.04)</td>
<td></td>
</tr>
<tr>
<td>TEXP4</td>
<td>.14</td>
<td>.04</td>
<td>3.23</td>
<td>(.05, .22)</td>
<td></td>
</tr>
<tr>
<td>EDLEVEL</td>
<td>.03</td>
<td>.04</td>
<td>.77</td>
<td>(−.05, .11)</td>
<td></td>
</tr>
<tr>
<td>ENROL</td>
<td>.30</td>
<td>.04</td>
<td>7.60</td>
<td>(.22, .37)</td>
<td></td>
</tr>
<tr>
<td>SAFE</td>
<td>−.16</td>
<td>.04</td>
<td>−3.71</td>
<td>(−.24, −.07)</td>
<td></td>
</tr>
</tbody>
</table>
PIRLS Data Analysis Example

Vibration of Effects Analysis

Vibration of Effects on zREAD
3600 regressions

\[ GIC_2 = \frac{\hat{y} - X\hat{\beta}}{\sigma} \]

- \( \hat{y} \) is the predicted value.
- \( X \) is the design matrix.
- \( \hat{\beta} \) is the estimated regression coefficient.
- \( \sigma \) is the standard error of the estimate.

**Legend:**
- \( \diamond \) An Effect Size, \( ES = \beta / \sqrt{\text{HC}} \), heteroscedastic-consistent (HC).
- \( \bullet \) A random biased ES.
- Red: Median GIC (and 95% CI) for a ES significance level.

**Axes:**
- Z: TEXP4 Effect Size
- 0 to 5

**Colors:**
- Purple: \( GIC_2 \) values
- Green: \( \hat{y} - X\hat{\beta} \) values
- Blue: \( \sigma \) values

**Scatter Plots:**
- Points represent the distribution of effect sizes across different conditions.
- Each condition is labeled with significance levels.

**Statistics:**
- DP precision \( \alpha \)
- # covariates in regression

**Significance Levels:**
- Significant at 0.05 level.
- Non-significant at 0.05 level.
- Strong significance at 0.005 level.

**Note:**
- The analysis is based on 3600 regressions to ensure robustness.
- The GIC2 statistic is used to evaluate the model fit.
- The effect size is adjusted for heteroscedasticity to provide accurate estimates.

**Interpretation:**
- The visual representation helps in understanding the variability and significance of the effect sizes across different conditions.
- The scatter plots provide a visual summary of the data, allowing for quick identification of significant patterns.

**Conclusion:**
- The analysis reveals significant effects at the 0.05 level, indicating that the model fit is robust and reliable.
Conclusions

• First study to show that the OLS estimator, as a posterior functional of the MDP, has (approximately) posterior mean given by WLS, and posterior covariance given by a weighted heteroscedastic-consistent sandwich covariance estimator.

• These posterior quantities (the WLS and sandwich estimators) are analytically manageable and permit fast computations for large data sets, and correspond to ridge regression (shrinkage) estimators for a particular choice of baseline distribution in the MDP.

• Illustrated MDP functional methodology through the analysis of simulated and real data (including VoE analyses).

• Manuscript:  http://arxiv.org/abs/1602.05155

• MDP methodology can be run using the “VoE analysis” menu option in my Bayesian Regression software: http://georgek.people.uic.edu/BayesSoftware.html

• In principle, the MDP functional methodology can be extended to processes of other Gibbs-type priors, such as the NIG process.

• For each of these other processes, the process variance is not available in closed form, which would then make Monte Carlo simulation necessary for posterior inference, and would make it more challenging to correspond the mean and variance of the process with its Polya urn scheme (posterior predictive) used for bootstrapping.