Drift vs Shift: Decoupling Trends & Changepoint Analysis

David S. Matteson, with Haoxuan Peter Wu & Sean Ryan
Cornell University (TRIPODS w/URochester) & the National Institute of Statistical Sciences (NISS)

Introduction

- **Goal**: Distinguish global/macro patterns from local/micro fluctuations
- 'Drift' describes the micro-level evolution of a process. This may appear as variation about gradual trends.
- 'Shifts' refer to discontinuities, rapid changes, or major breaks in trend. These represent macro-level changes in a process.
- Both might be mechanistically or stochastically generated and/or modeled. However, causes of shifts are typically different from those of drift.
- While understanding such differences is a prime objective, this first requires distinguishing: **Drift vs Shift**.

Tools include:

- Trend Filtering
- Stochastic Volatility
- Outlier Detection
- Dynamic/Adaptive Shrinkage
- Dynamic Linear Models (DLM)
- Change Point Analysis
- Bayesian (Time Series) Analysis
- Machine Learning (Regularization)

Challenges

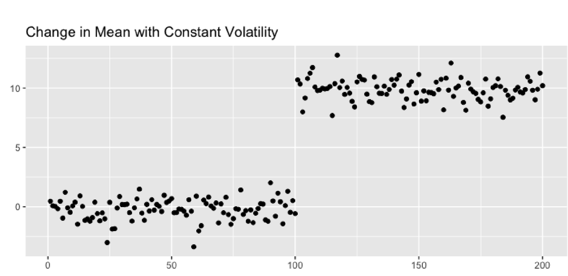


Fig 1 Change in Mean

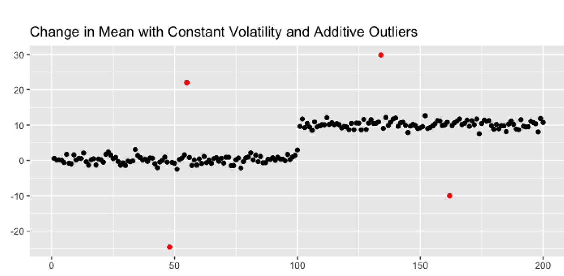


Fig 2 Adaptive Outliers

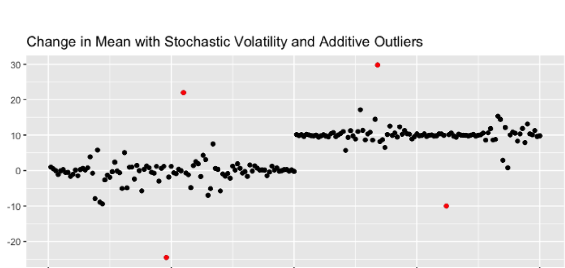


Fig 3 Stochastic Volatility

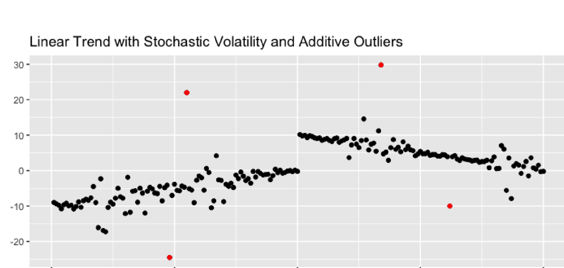


Fig 4 Linear Trend

- **Outliers** violate common Gaussian noise assumptions.
- **Heterogeneity** leads to over-prediction of changepoints.

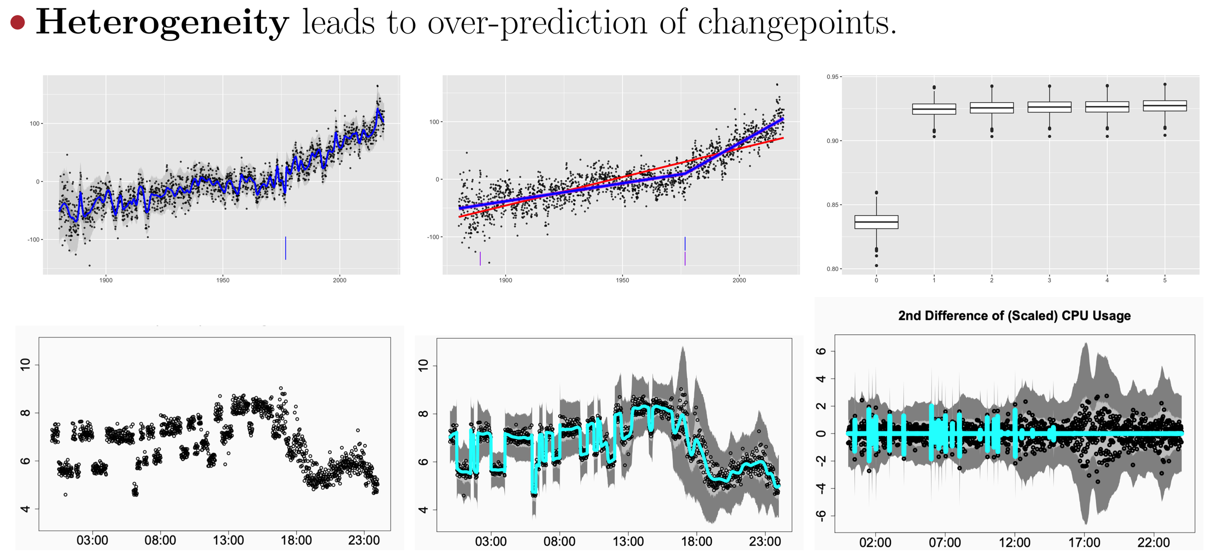


Fig 5 Global land surface air temperature (top); CPU cloud usage (bottom).

- Real world data has complex patterns and trends.
- Outliers and heterogeneity are the norm.
- Nature of changepoints ambiguous.

Solutions

- Model based ABCO: Adaptive Bayesian Changepoints w/ Outliers[1].
- A two-step Bayesian 'decoupling' method developed via DLM[2].

ABCO Model

Given a time series $\{y_t\}$, ABCO supposes the decomposition:

$$y_t = \underbrace{\beta_t}_{\text{mean signal}} + \underbrace{\zeta_t}_{\text{additive outlier}} + \underbrace{\epsilon_t}_{\text{heteroskedastic noise}}$$

- **Trend Signal $\{\beta_t\}$**
  Ref "Dynamic Shrinkage Process" [3], ABCO uses global-local shrinkage priors on the Dth order difference ($\triangle^D$, $D = 1, 2$) on the state variable $\{\beta_t\}$:

$$\triangle^D \beta_t = \omega_t, \qquad \omega_t \sim N(0, \tau_\omega^2 \lambda_{\omega,t}^2 = e^{h_t}),$$
$$h_{t+1} = \mu + (\phi_1 + \phi_2 s_t)(h_t - \mu) + \eta_{t+1}, \qquad \eta_{t+1} \sim Z(\alpha, \beta, 0, 1).$$

  Z-distribution: log inverted Beta; heavy left-tail

  - **Changepoint:** threshold γ & indicator $s_t = \begin{cases} 1 & \text{if } \log(\omega_t^2) > \gamma \\ 0 & \text{if } \log(\omega_t^2) \leq \gamma \end{cases}$.

- **Additive Outlier $\{\zeta_t\}$**
  The outlier term $\{\zeta_t\}$ follows a "horseshoe+" shrinkage prior:

$$(\zeta_t | \sigma_{\zeta,t}) \sim N(0, \sigma_{\zeta,t}^2)$$
$$(\sigma_{\zeta,t} | \tau_\zeta, \eta_{\zeta,t}) \sim C^+(0, \tau_\zeta \eta_{\zeta,t})$$
$$\tau_\zeta \sim C^+(0, \sigma_{\tau,\zeta})$$
$$\eta_{\zeta,t} \sim C^+(0, \sigma_{\eta,\zeta})$$

  with half-Cauchy $C^+(\cdot)$ and prior shrinkage hyper-parameters $\sigma_{\tau,\zeta}, \sigma_{\eta,\zeta}$.

  - **Outlier:** custom cutoff & locally adaptive score $o_t := \widetilde{E}\left(\frac{\sigma_{\zeta,t}^2}{\sigma_{\zeta,t}^2 + \sigma_{\epsilon,t}^2}\right)$.

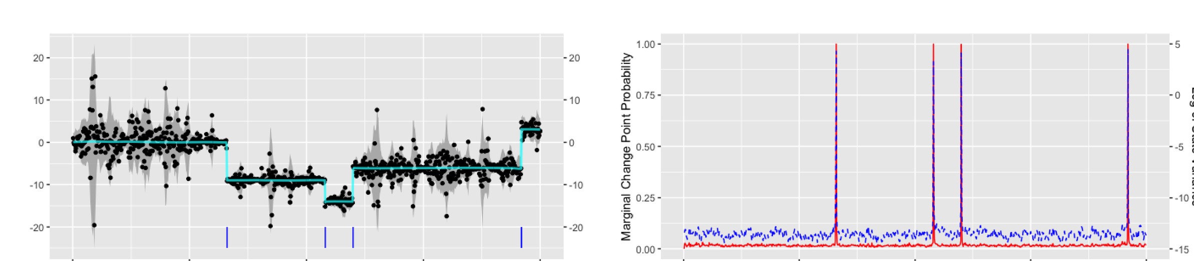
- **Heteroskedastic Noise $\{\epsilon_t, \sigma_{\epsilon,t}^2\}$**
  The noise $\{\sigma_{\epsilon,t}^2\}$ follows a stochastic volatility model of order 1.

$$y_t = \beta_t + \zeta_t + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_{\epsilon,t}^2),$$
$$\log(\sigma_{\epsilon,t}^2) = \mu_\epsilon + \phi_\epsilon(\log(\sigma_{\epsilon,t}^2) - \mu_\epsilon) + \xi_{\epsilon,t}, \qquad \xi_{\epsilon,t} \sim N(0, \sigma_\xi^2).$$

- **Dynamic Regression Generalizations**
  Set $\boldsymbol{x}_t = (x_{1,t}, ..., x_{p,t})$ as p predictors at time t, and $\boldsymbol{\omega}, \boldsymbol{h}, \boldsymbol{\mu}, \boldsymbol{\phi}, \boldsymbol{\eta}$ analogously.

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \zeta_t + \epsilon_t \qquad \triangle^D \beta_{t+1} = \boldsymbol{\omega}_t$$
$$\omega_{j,t} \sim N(0, \tau_{\omega,0}^2 \tau_{\omega,j}^2 \lambda_{\omega,j,t}^2 = e^{h_{j,t}}) \qquad \boldsymbol{h}_{t+1} = \boldsymbol{\mu} + (\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2 \boldsymbol{s}_t)(\boldsymbol{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_{t+1}$$

ABCO Simulations



Algorithms	Rand Avg. ↑	Adj. Rand Avg. ↑	Avg. Diff. CP ↓	Avg. Dist. to True ↓
ABCO	**0.938**	**0.873**$_{(0.022)}$	**0.51**	**0.03**
E.Divisive	0.848	0.655$_{(0.022)}$	3.68	58.45
Pelt	0.730	0.361$_{(0.016)}$	10.88	128.36
WBS	0.686	0.352$_{(0.021)}$	106.65	144.45
BCP	0.841	0.595$_{(0.024)}$	22.00	60.16

Fig 6 Length-1000 time series with shift and SV(1) variance model
$\log(\sigma_{\epsilon,t}^2) = \phi_\epsilon \log(\sigma_{\epsilon,t-1}^2) + \alpha_t, \quad \alpha_t \sim N(0, \sigma_\alpha^2), \quad \phi_\epsilon = 0.9, \quad \sigma_\alpha = 0.4.$

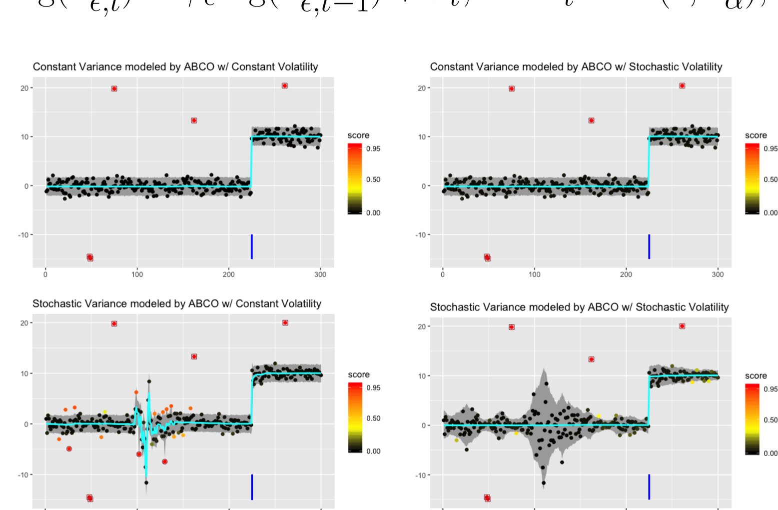


Fig 7 Robustness plus Outlier Scoring.

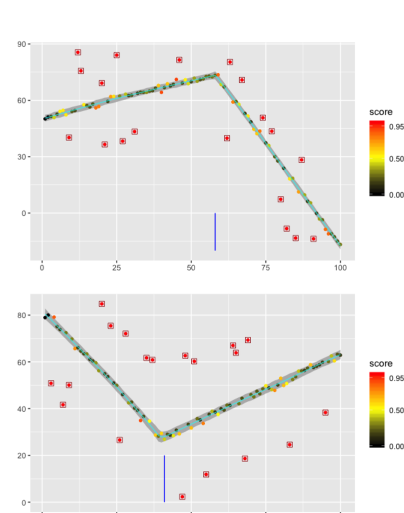


Fig 8 Linear Meetup Model.

ABCO Applications

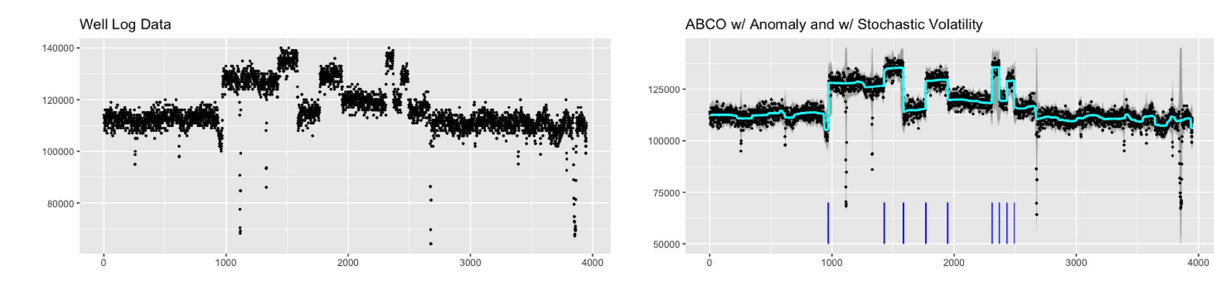


Fig 9 On Well-Log data, nuclear magnetic response within rock formations, originally published in [4] as a good framework for changepoint detection.

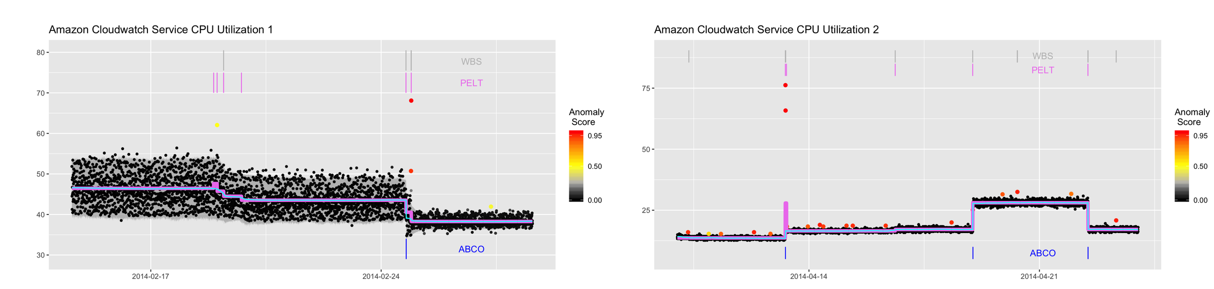


Fig 10 On Amazon Cloudwatch Service CPU Utilization data.

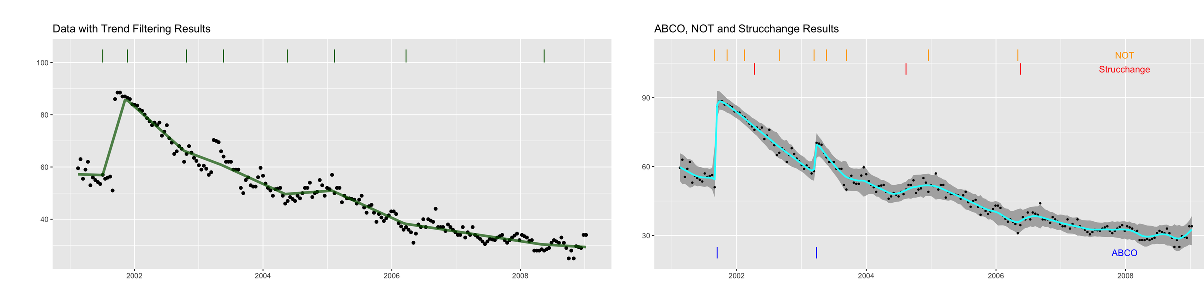


Fig 11 On George W. Bush Approval Rating data.

Decoupling Approach

- **Dynamic Linear Models (DLM)**
  Given a time series $\boldsymbol{Y} = (y_1, ..., y_n)'$, a predictor series $\boldsymbol{X} = (x_1, ..., x_n)'$,

$$y_t = x_t \beta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{\epsilon,t}^2),$$
$$\triangle^D \beta_t = \omega_t, \qquad \omega_t \sim N(0, \sigma_\omega^2).$$

- **Decoupled Regularized Loss**
  Denote $\bar{\boldsymbol{\beta}}$ as the posterior mean of k MCMC draws $\{\boldsymbol{\beta}^{(i)}, i = 1, ..., k\}$ of $\{\beta_t\}$.

$$\textbf{Decoupled loss: } L_\lambda(\widetilde{\boldsymbol{\beta}}) = ||\boldsymbol{W}^{1/2}(\boldsymbol{X} \circ \bar{\boldsymbol{\beta}} - \boldsymbol{X} \circ \widetilde{\boldsymbol{\beta}})||_2^2 + q_\lambda(\widetilde{\boldsymbol{\beta}}).$$

  - $\boldsymbol{W} = \text{diag}(w_1, ..., w_n)$ is diagonal with weights for each measurement being
$$w_i = 1/\bar{\sigma}_{\epsilon,i}^2, \text{ for } i = 1, ..., n.$$
  - Penalty function $q_\lambda()$ induces sparsity into $\tilde{\boldsymbol{\beta}}$ with form
$$q_\lambda(\tilde{\boldsymbol{\beta}}) = \lambda \sum_t \frac{1}{|\psi_t|} |\triangle^D \beta_t|,$$

  where $\psi_t = \frac{1}{k}\sum_{i=1}^k \triangle^D \beta_t^{(i)}$ and $D = 1, 2$ controls the type of change.

- **Changepoint Selection**
  Given λ, denote η_λ as the time indices which $\{\triangle^D \tilde{\beta}_t \neq 0\}$.

$$\textbf{Diagnostic tool: } R_\lambda^2 = \frac{1}{k}\sum_{i=1}^k \frac{||\boldsymbol{\beta}^{(i)} - \boldsymbol{\beta}_\eta^{(i)}||^2}{||\boldsymbol{\beta}^{(i)} - \bar{\boldsymbol{\beta}}^{(i)}||^2}$$

  where $\bar{\boldsymbol{\beta}}^{(i)} = \frac{1}{n}\sum_{t=1}^n \beta_t^{(i)}$, and the optimal λ determined by least changepoints given $E[R_\lambda^2]$ exceeds a certain threshold.

* **Multiple Predictors & Covariates**
  Set predictors $\boldsymbol{X} = \text{blockdiag}(\boldsymbol{x}_1', ..., \boldsymbol{x}_n')$ with $\boldsymbol{x}_i = (x_{i,1}, ..., x_{i,p})'$, and covariates $\boldsymbol{Z} = (\boldsymbol{z}_1, ..., \boldsymbol{z}_n)$ with $\boldsymbol{z}_i = (z_{i,1}, ..., z_{i,l})'$.

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + \boldsymbol{\alpha}'\boldsymbol{z}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_{\epsilon,t}^2), \quad \triangle^D \boldsymbol{\beta}_t = \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \Sigma_{\omega,t}).$$

  Extend the model with the decoupled loss:

$$L_\lambda(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}) = ||\boldsymbol{W}^{1/2}(\boldsymbol{X}\bar{\boldsymbol{\beta}} + \boldsymbol{Z}\bar{\boldsymbol{\alpha}} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} - \boldsymbol{Z}\widetilde{\boldsymbol{\alpha}})||_2^2 + q_\lambda(\widetilde{\boldsymbol{\beta}}),$$
$$q_\lambda(\widetilde{\boldsymbol{\beta}}) = \lambda \sum_{t=1}^n \sum_{g=1}^G \frac{1}{|\boldsymbol{\psi}_{g,t}|} |\triangle^D \boldsymbol{\beta}_{g,t}|.$$

Decoupling Simulations

Bayesian DLM

Decoupled

Comp. F1-Score

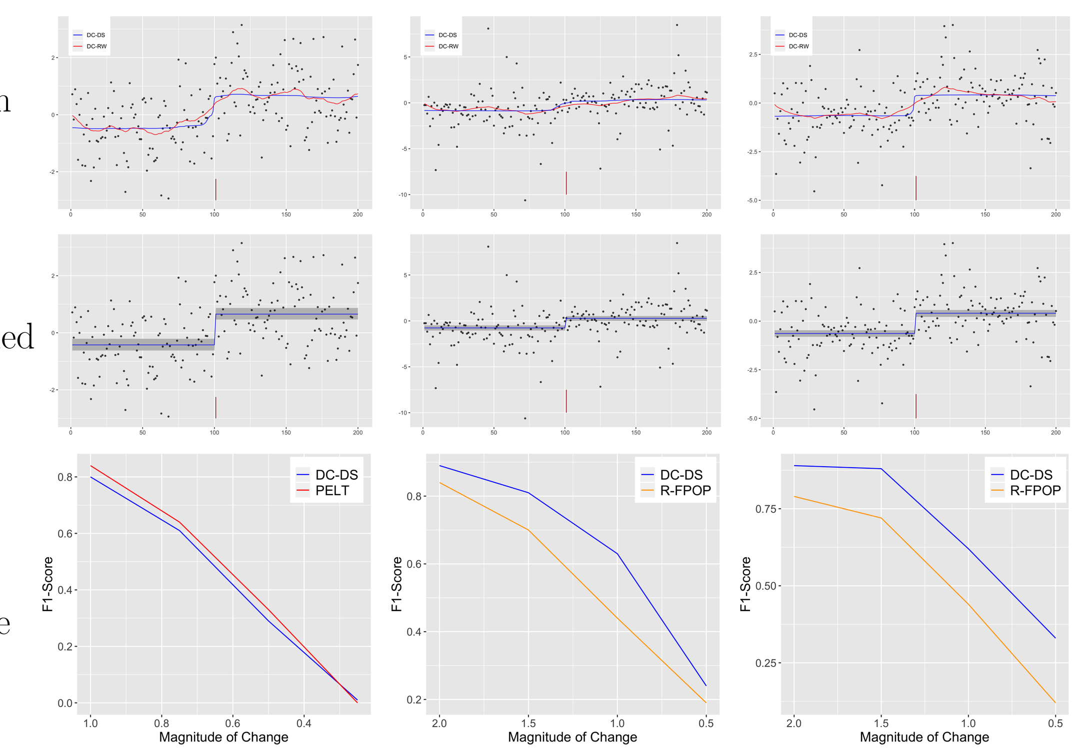


Fig 12 Gaussian Noise (left); Outliers (middle); Stochastic Volatility (right).
DC-DS: decoupled results with shrinkage. DC-RW: decoupled with random walk.

Decoupling Applications

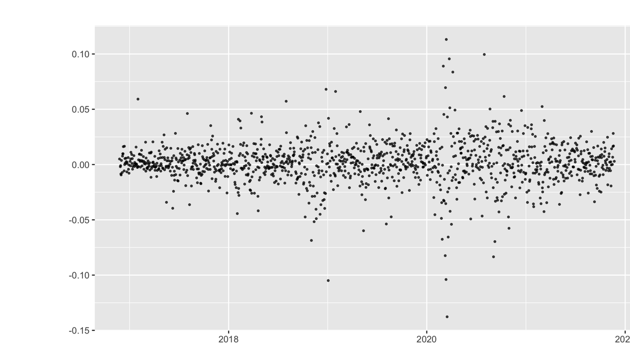


Fig 13 Apple Daily Stock Return

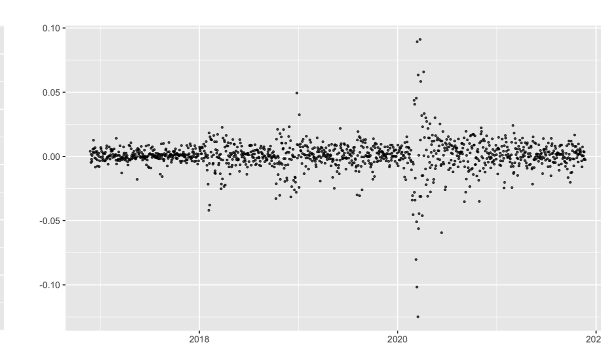


Fig 14 SP500 Daily Stock Return

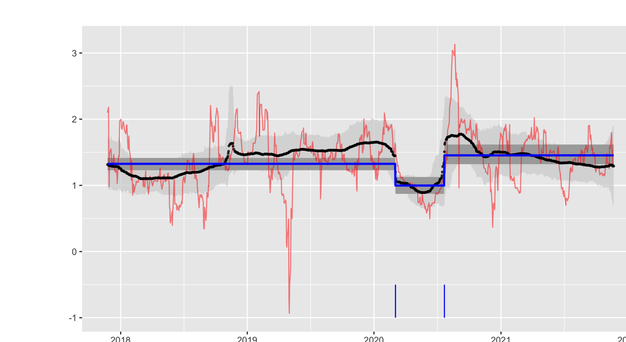


Fig 15 Decoupled vs Rolling OLS $\{\beta_t\}$

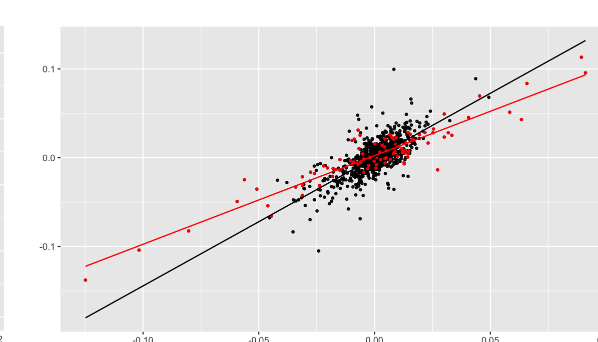


Fig 16 Scatter of Paired Returns

Conclusions

- A framework for inferring changepoints from posteriors produced by Bayesian time-varying parameter models.
- By decoupling trend modeling and changepoint analysis, we allow fitting an arbitrarily complex model to deal with intricacies inherent in data.
- Extensions: higher order trend changes, regression coefficients, multivariate.

References

[1] Haoxuan Wu and David S Matteson. Adaptive bayesian changepoint analysis and local outlier scoring. arXiv preprint arXiv:2011.09437, 2020.

[2] Haoxuan Wu, Sean Ryan, and David S Matteson. Decoupling trends and changepoint analysis. arXiv preprint arXiv:2201.06606, 2022.

[3] Daniel Kowal, David S. Matteson, and David Ruppert. Dynamic shrinkage process. Journal of the Royal Statistical Society: Series B, 2018.

[4] Joseph J. K. Ó Ruanaidh and William J. Fitzgerald. Numerical bayesian methods applied to signal processing. Statistics and Computing, 1996.

[5] Sangjoon Kim, Neil Shephard, and Siddhartha Chib. Stochastic volatility: likelihood inference and comparison with arch models. Review of Economic Studies, 65:361–393, 1998.

[6] David S. Matteson and Nicholas A. James. A nonparametric approach for multiple change point analysis of multivariate data. Journal of the American Statistical Association, 109:334–345, 2014.

[7] Anindya Bhadra, Jyotishka Datta, Nicholas G. Polson, and Brandon Willard. The horseshoe+ estimator for ultra sparse signals. Bayesian Analysis, 12:1105–1131, 2017.

[8] Piotr Fryzlewicz. Wild binary segmentation for multiple change-point detection. Annals of Statistics, 42:2243–2281, 2014.

[9] Chandra Erdman and John W. Emerson. A fast bayesian change point analysis for the segmentation of microarray data. Bioinformatics, 24:2143–2148, 2008.