Empirical Bayes Derivative Estimates

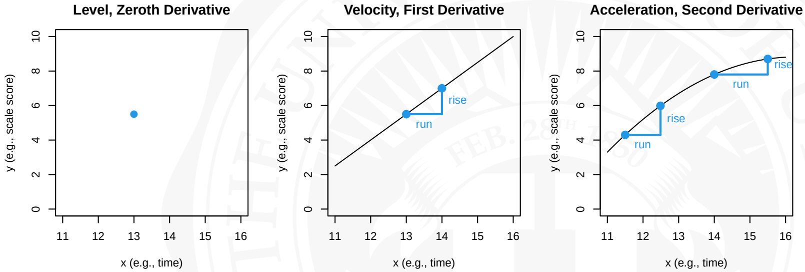
Pascal R. Deboeck
University of Utah

Modern Modeling Methods 2023
University of Connecticut
June 27, 2023

Introduction

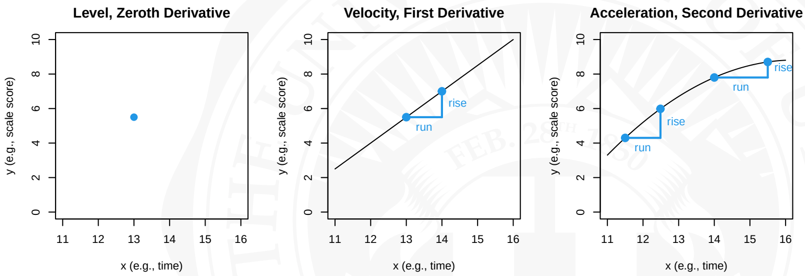
- Derivative Estimation, Empirical Bayes Derivative Estimates
- Simulations: Short, Long Time Series
- Substantive Examples

Derivatives



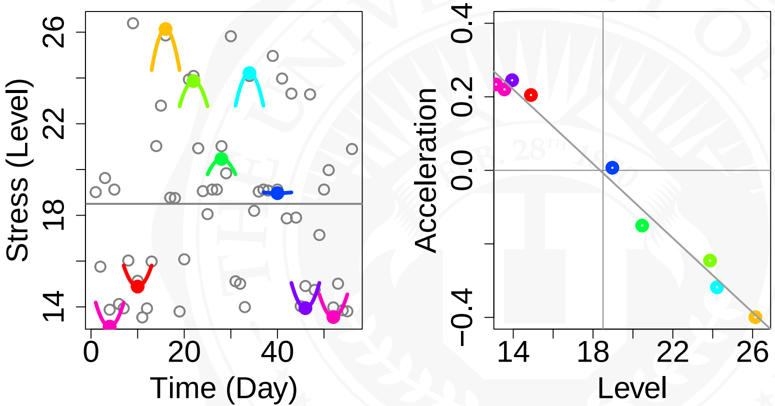
- Zeroth: x, First: dx/dt, Second: d^2x/dt^2

Derivatives



- Zeroth: x, First: dx/dt, Second: d^2x/dt^2
- Relations between derivatives

Differential Equation Modeling



Order-Plus-One

- Difference Scores $(x_t - x_{t-1}, x_{post} - x_{pre})$
  - Zeroth Difference: x_t
  - First Difference: $x_{t+\Delta} - x_t$
  - Second Difference: $(x_{t+\Delta} - x_t) - (x_t - x_{t-\Delta})$

[1] Boker & Nesselroade, 2002

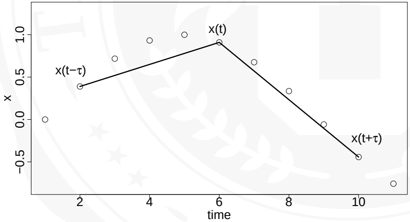
Order-Plus-One

- Difference Scores $(x_t - x_{t-1}, x_{post} - x_{pre})$
  - Zeroth Difference: x_t
  - First Difference: $x_{t+\Delta} - x_t$
  - Second Difference: $(x_{t+\Delta} - x_t) - (x_t - x_{t-\Delta})$
- Local Linear Approximation[1]
  - Zeroth Derivative: x_t
  - First Derivative: $(x_{t+\Delta} - x_t)/\tau\Delta$
  - Second Derivative: $(x_{t+\Delta} + x_{t-\Delta} - 2x_t)/\tau^2\Delta^2$

[1] Boker & Nesselroade, 2002

Order-Plus-One

- Difference Scores $(x_t - x_{t-1}, x_{post} - x_{pre})$
  - Zeroth Difference: x_t
  - First Difference: $x_{t+\Delta} - x_t$
  - Second Difference: $(x_{t+\Delta} - x_t) - (x_t - x_{t-\Delta})$
- Local Linear Approximation[1]
  - Zeroth Derivative: x_t
  - First Derivative: $(x_{t+\Delta} - x_t)/\tau\Delta$
  - Second Derivative: $(x_{t+\Delta} + x_{t-\Delta} - 2x_t)/\tau^2\Delta^2$
  - Time explicit, τ smoothing



[1] Boker & Nesselroade, 2002

Regression Methods

- Generalized Local Linear Approximation (GLLA)[2]
  - $X = \{X_t, X_{t+1}, X_{t+2}, X_{t+3}, X_{t+4}\}$
  - $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0, β_1, β_2

[2] Boker, Deboeck, Edler & Keel, 2009
[3] Deboeck, 2010

Regression Methods

- Generalized Local Linear Approximation (GLLA)[2]
  - $X = \{X_t, X_{t+1}, X_{t+2}, X_{t+3}, X_{t+4}\}$
  - $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0, β_1, β_2
  - Non-interval sampling, any order of derivative, measurement error

[2] Boker, Deboeck, Edler & Keel, 2009
[3] Deboeck, 2010

Regression Methods

- Generalized Local Linear Approximation (GLLA)[2]
  - $X = \{X_t, X_{t+1}, X_{t+2}, X_{t+3}, X_{t+4}\}$
  - $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0, β_1, β_2
  - Non-interval sampling, any order of derivative, measurement error
- Generalized Orthogonal Local Derivative Estimates (GOLD)[3]
  - EQ1: $X = \beta_0 + \varepsilon$
  - EQ2: $X = \beta_0 + \beta_1(Time) + \varepsilon$
  - EQ3: $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0 (EQ1), β_1 (EQ2), β_2 (EQ3)

---

[2] Boker, Deboeck, Edler & Keel, 2009

[3] Deboeck, 2010

Regression Methods

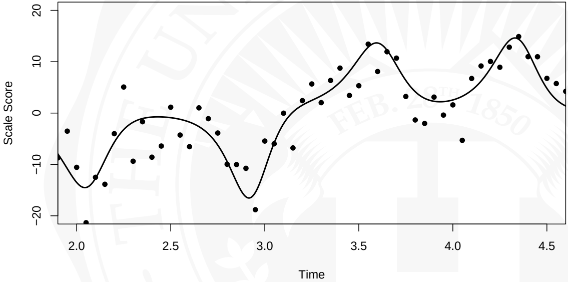
- Generalized Local Linear Approximation (GLLA)[2]
  - $X = \{X_t, X_{t+1}, X_{t+2}, X_{t+3}, X_{t+4}\}$
  - $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0, β_1, β_2
  - Non-interval sampling, any order of derivative, measurement error
- Generalized Orthogonal Local Derivative Estimates (GOLD)[3]
  - EQ1: $X = \beta_0 + \varepsilon$
  - EQ2: $X = \beta_0 + \beta_1(Time) + \varepsilon$
  - EQ3: $X = \beta_0 + \beta_1(Time) + \beta_2(\frac{1}{2}Time^2) + \varepsilon$
  - Derivatives: β_0 (EQ1), β_1 (EQ2), β_2 (EQ3)
  - Advantages are the same as GLLA, orthogonal partitioning of variance

[2] Boker, Deboeck, Edler & Keel, 2009

[3] Deboeck, 2010

Non-independent Estimation

- Functional Data Analysis[4]
- Repeated observations: complex, but smooth, process



- Polynomial segments (like GLLA), but time-adjacent polynomials the derivatives are constrained to be continuous (to specified derivative)

[4] Ramsay & Silverman, 2002, 2005

Non-independent Estimation

- Empirical Bayes Derivative Estimates

Non-independent Estimation

- Empirical Bayes Derivative Estimates
- Panel Data (large n, few t), Developmental Psychology

Non-independent Estimation

- Empirical Bayes Derivative Estimates
- Panel Data (large n, few t), Developmental Psychology
- Multilevel Model, random effects for Intercepts and Slopes

Non-independent Estimation

- Empirical Bayes Derivative Estimates
- Panel Data (large n, few t), Developmental Psychology
- Multilevel Model, random effects for Intercepts and Slopes
- Derivatives (Level, Velocity, Acceleration) as random effects

Short Time Series Data Structure

$id(i)$	X_{it}	$Time$	$\frac{1}{2}Time^2$
1	X_{11}	-1.5	1.125
1	X_{12}	-0.5	0.125
1	X_{13}	0.5	0.125
1	X_{14}	1.5	1.125
2	X_{21}	-1.5	1.125
2	X_{22}	-0.5	0.125
2	X_{23}	0.5	0.125
2	X_{24}	1.5	1.125
⋮	⋮	⋮	⋮

Long Time Series Data Structure

id	$Windows(i)$	X_{it}	$Time$	$\frac{1}{2} Time^2$
1	1	X_{11}	-1.5	1.125
1	1	X_{12}	-0.5	0.125
1	1	X_{13}	0.5	0.125
1	1	X_{14}	1.5	1.125
1	2	X_{12}	-1.5	1.125
1	2	X_{13}	-0.5	0.125
1	2	X_{14}	0.5	0.125
1	2	X_{15}	1.5	1.125
1	3	X_{13}	-1.5	1.125
1	3	X_{14}	-0.5	0.125
1	3	X_{15}	0.5	0.125
1	3	X_{16}	1.5	1.125
⋮	⋮	⋮	⋮	⋮

Empirical Bayes Derivative Estimates

$$X_{it} = \beta_{0i} + \beta_{1i}(Time) + \beta_{2i}(\frac{1}{2}Time^2) + \varepsilon_{it}$$
$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$
$$\beta_{2i} = \gamma_{20} + u_{2i}$$

- Multivariate normal distribution of derivatives: β_{0i}, β_{1i}, β_{2i}

Empirical Bayes Derivative Estimates

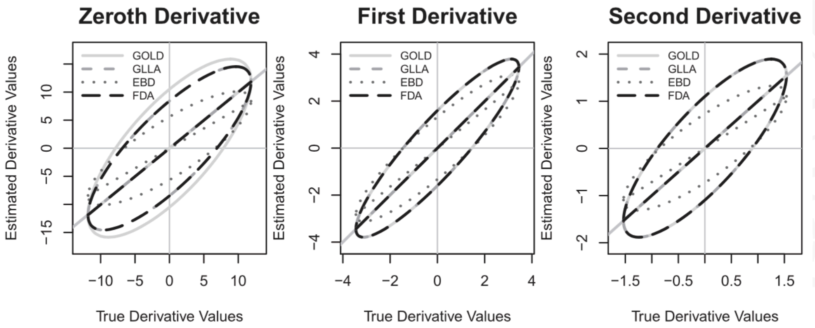
$$X_{it} = \beta_{0i} + \beta_{1i}(Time) + \beta_{2i}(\tfrac{1}{2}Time^2) + \varepsilon_{it}$$
$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$
$$\beta_{2i} = \gamma_{20} + u_{2i}$$

- Multivariate normal distribution of derivatives: β_{0i}, β_{1i}, β_{2i}
- Subsequent estimation of random effects: "Conditional mean/mode" estimates
- Application of Bayes' formula with model estimates as known prior, data for each "group"

Simulations

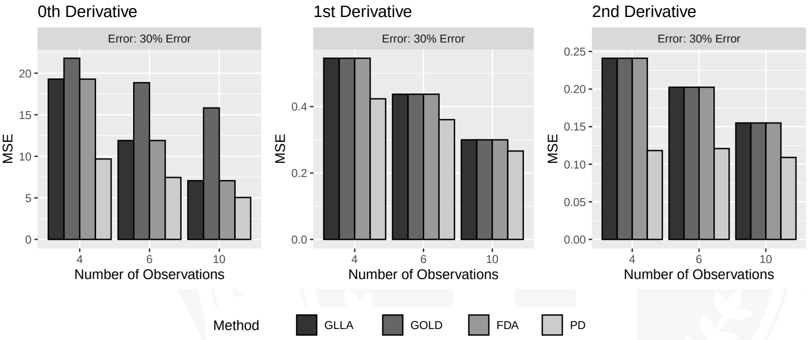
- Two Simulations: Short, Long time series
- GLLA, GOLD, FDA, EBD
- Short Time Series
  - Quadratic time series
  - 4-10 observations per time series
  - 25, 50, 100, 500 time series (i.e., many individuals)
  - Measurement error: 10%, 30%, 50%

Results



- 100 time series, 30% measurement error, 6 observations/time series

Results, MSE

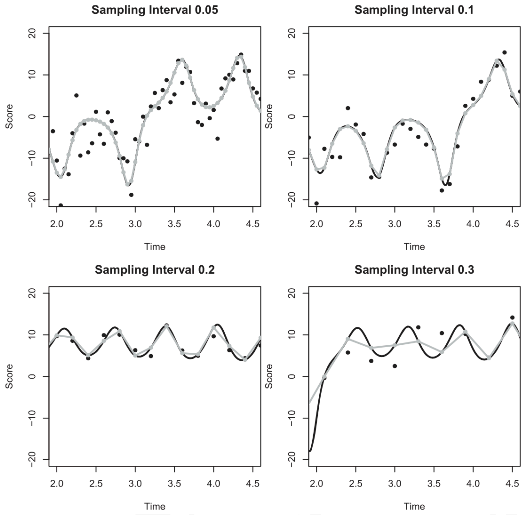


- 100 time series, 30% measurement error

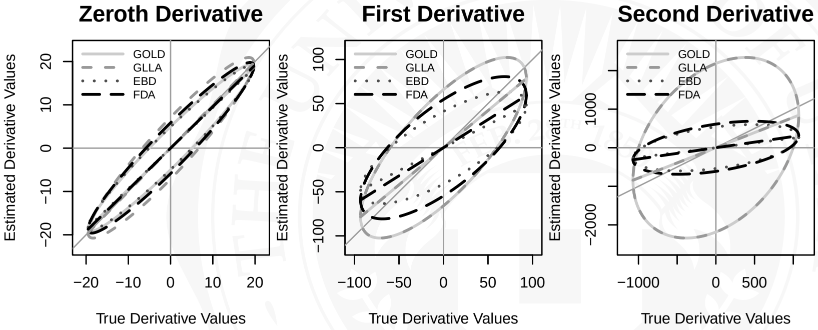
Long Time Series Simulation

- “x” component of Lorenz Attractor
- Varied sampling rate (0.05, 0.10, 0.20, 0.30)
- Embedding 5
- Time series of 25, 50, 100, 500 observations
- Measurement error: 10%, 30%, 50%

Long Time Series, Sampling Rates

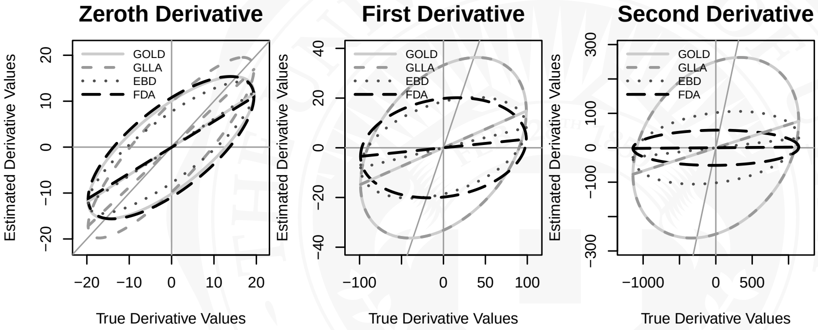


Results: High Fidelity, Sampling Rate 0.05



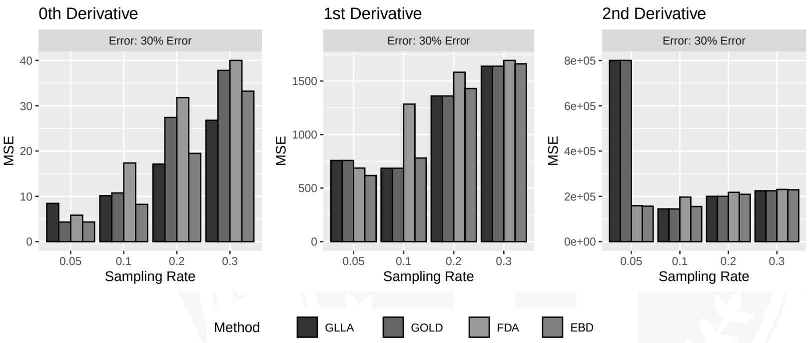
- Time series length of 100, 30% error

Results: Low Fidelity, Sampling Rate 0.20



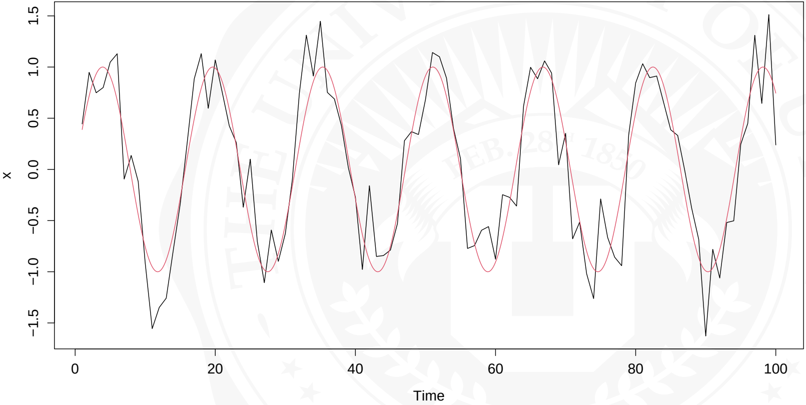
- Time series length of 100, 30% error

Results, MSE

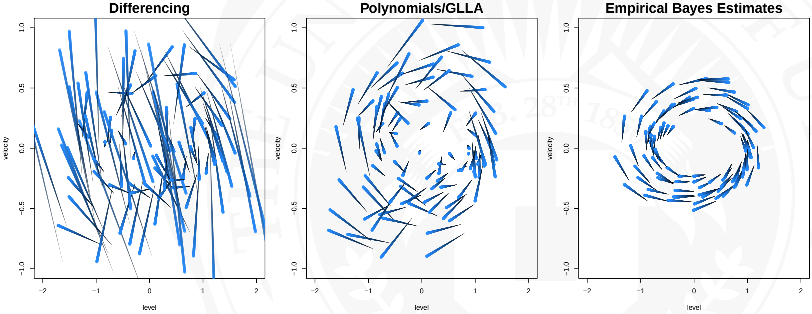


- Time series length of 100, 30% error

Differencing — GLLA — EBD



Differencing — GLLA — EBD



Limitations

- Assumption of Multivariate Normality

Limitations

- Assumption of Multivariate Normality
- Fairly robust to non-normality of random effects

Limitations

- Assumption of Multivariate Normality

- Fairly robust to non-normality of random effects
- Same mean/covariance matrix across all estimates
  - People with Different Dynamics
  - Non-stationary Time Series

Limitations

- Assumption of Multivariate Normality

- Fairly robust to non-normality of random effects
- Same mean/covariance matrix across all estimates
  - People with Different Dynamics
  - Non-stationary Time Series
- Linear relations between derivatives

Variations

- Multilevel Modeling Advantages

Variations

- Multilevel Modeling Advantages
- Unequal Observation Intervals/Missing Data

Variations

- Multilevel Modeling Advantages
- Unequal Observation Intervals/Missing Data
- Generalized Linear Mixed Models (Categorical, Poisson data)

Variations

- Multilevel Modeling Advantages

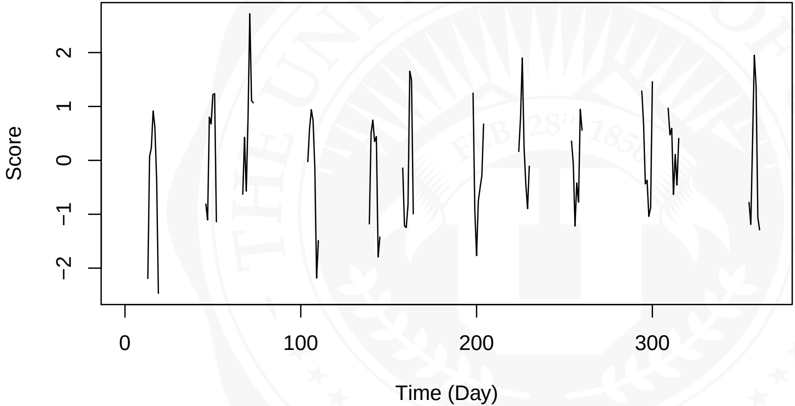
- Unequal Observation Intervals/Missing Data
- Generalized Linear Mixed Models (Categorical, Poisson data)
- Simultaneously Detrend Data

Variations

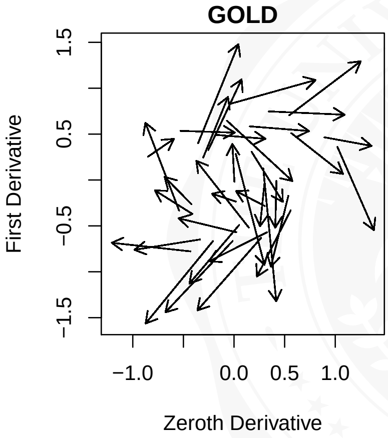
- Multilevel Modeling Advantages

- Unequal Observation Intervals/Missing Data
- Generalized Linear Mixed Models (Categorical, Poisson data)
- Simultaneously Detrend Data
- New intensive sampling schemes

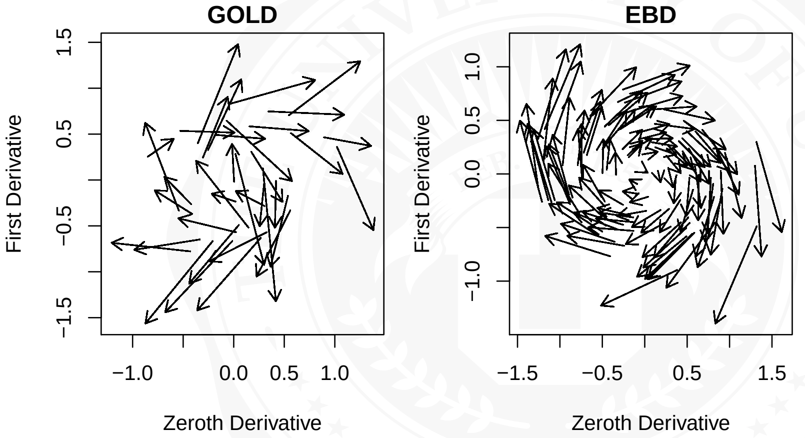
7/30 Sampling Scheme



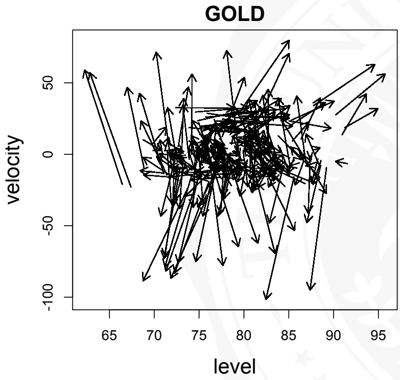
7/30 Sampling Scheme



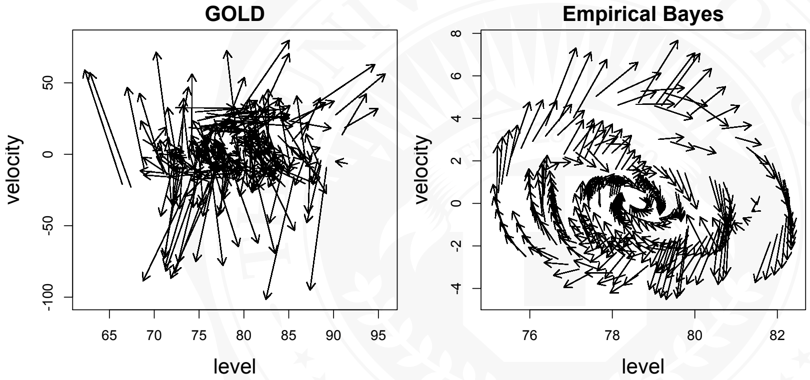
7/30 Sampling Scheme



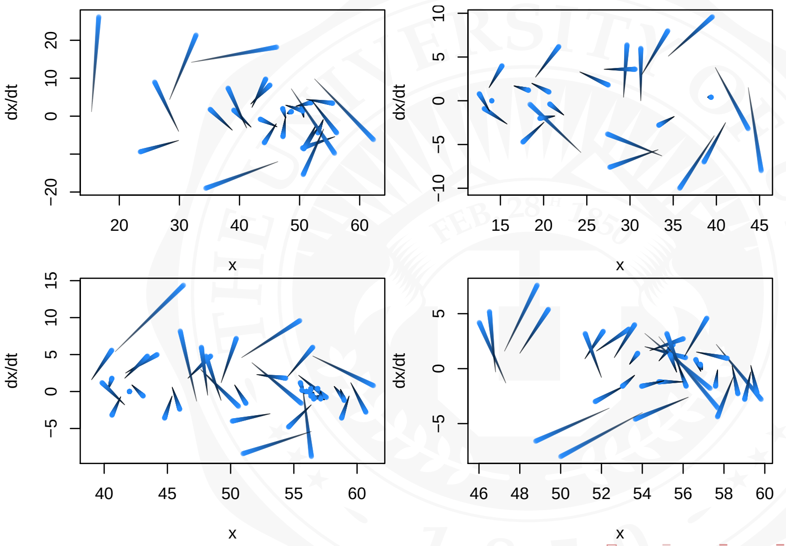
Heart Rate Data



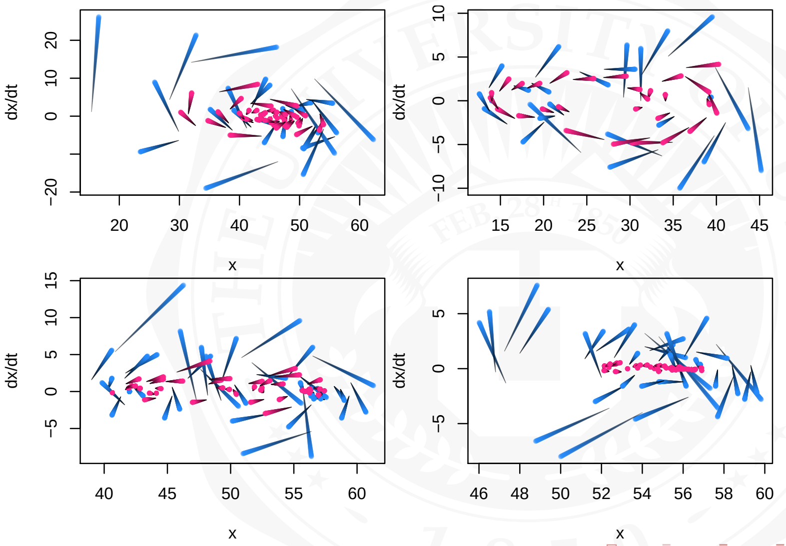
Heart Rate Data



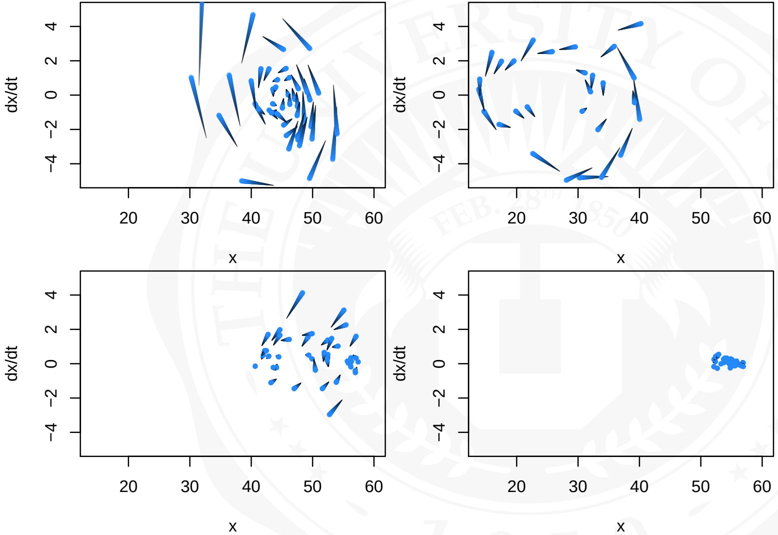
Positive Affect, Daily for 56 Days (GOLD)



Positive Affect, Daily for 56 Days (GOLD, EBD)



Positive Affect, Daily for 56 Days (EBD)



Conclusion

- Derivative Estimation
  - Order-Plus-One (Differences, LLA), Regression Methods (GLLA, GOLD), Non-independent Estimation (FDA)
  - Empirical Bayes Derivative Estimates (Multilevel Models, random effects)
  - Simulations highlight attenuation, greatly reduced variance
  - Generally as efficient or better than GLLA/GOLD/FDA
- Variations
  - Unequal Observation Intervals/Missing Data
  - Generalized Linear Mixed Models (Categorical, Poisson data)
  - Simultaneously Detrend Data
  - New intensive sampling schemes (7/30 for a year)
- Substantive: Heart Rate, Positive Affect

- Paper: Deboeck, P. R. (2020). Empirical Bayes Derivative Estimates. Multivariate Behavioral Research, 55:3, 382–404.

- Substantive State-Space Modeling: Montpetit, M. A., Bergeman, C. S., Deboeck, P. R., Tiberio, S. S. & Boker, S. M. (2010) Resilience-As-Process: Negative Affect, Stress, and Coupled Dynamical Systems. Psychology and Aging, 25 (3), 631–640.
- GLLA: Boker, S. M., Deboeck, P. R., Edler, C. & Keel, P. K. (2009). Generalized Local Linear Approximation of Derivatives from Time Series. In S. Chow, E. Ferrer & F. Hsieh (Eds.), Statistical Methods for Modeling Human Dynamics: An Interdisciplinary Dialogue, pp. 161–178. New York, NY: Taylor & Francis Group.
- GOLD: Deboeck, P. R. (2010). Estimating Dynamical Systems, Derivative Estimation Hints from Sir Ronald A. Fisher. Multivariate Behavioral Research, 43 (4), 725–745.