Discovering psycholinguistic effect timecourses with deconvolutional time series regression

Cory Shain

November 7, 2018, *Department of Cognitive Science, Johns Hopkins University*

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
    - Deconvolutional time series regression (DTSR)
    - Continuous-time mixed-effects deconvolutional regression model
    - Can be applied to any time series
- **Results:**
    - Recovers known temporal structures with high fidelity
    - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

This talk in one slide

- Temporal diffusion of effects can be a serious confound in psycholinguistic data
- Modeling temporal diffusion is problematic with existing tools
- **Proposal:**
  - Deconvolutional time series regression (DTSR)
  - Continuous-time mixed-effects deconvolutional regression model
  - Can be applied to any time series
- **Results:**
  - Recovers known temporal structures with high fidelity
  - Finds plausible, replicable, and high-resolution estimates of temporal structure in reading data
- Documented open-source Python package supports easy adoption

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Time matters a lot in psycholinguistics
  - Psycholinguistic data are generated by people with brains
  - The brain is a dynamical system that responds to its environment in time
  - Most (all?) psycholinguistic data are underlyingly time series
  - The brain's response to a stimulus may be slow (temporally diffuse)
  - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Time matters a lot in psycholinguistics
  - Psycholinguistic data are generated by people with brains
  - The brain is a dynamical system that responds to its environment in time
  - Most (all?) psycholinguistic data are underlyingly time series
  - The brain's response to a stimulus may be slow (temporally diffuse)
  - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Time matters a lot in psycholinguistics
    - Psycholinguistic data are generated by people with brains
    - The brain is a dynamical system that responds to its environment in time
    - Most (all?) psycholinguistic data are underlyingly time series
    - The brain's response to a stimulus may be slow (temporally diffuse)
    - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Time matters a lot in psycholinguistics
  - Psycholinguistic data are generated by people with brains
  - The brain is a dynamical system that responds to its environment in time
  - Most (all?) psycholinguistic data are underlyingly time series
  - The brain's response to a stimulus may be slow (temporally diffuse)
  - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Time matters a lot in psycholinguistics
  - Psycholinguistic data are generated by people with brains
  - The brain is a dynamical system that responds to its environment in time
  - Most (all?) psycholinguistic data are underlyingly time series
  - The brain's response to a stimulus may be slow (temporally diffuse)
  - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Time matters a lot in psycholinguistics
  - Psycholinguistic data are generated by people with brains
  - The brain is a dynamical system that responds to its environment in time
  - Most (all?) psycholinguistic data are underlyingly time series
  - The brain's response to a stimulus may be slow (temporally diffuse)
  - Psycholinguistic measures may capture lingering response to preceding events

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Signal processing provides a framework for capturing temporal diffusion
  - Stimuli and responses can be recast as convolutionally-related signals
  - Relation described by an impulse response function (IRF)
  - If we can discover the structure of the IRF (deconvolution), we can convolve predictors with it to obtain a model of the response that takes diffusion directly into account

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Signal processing provides a framework for capturing temporal diffusion
  - Stimuli and responses can be recast as convolutionally-related signals
  - Relation described by an impulse response function (IRF)
  - If we can discover the structure of the IRF (deconvolution), we can convolve predictors with it to obtain a model of the response that takes diffusion directly into account

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Signal processing provides a framework for capturing temporal diffusion
  - Stimuli and responses can be recast as convolutionally-related signals
  - Relation described by an impulse response function (IRF)
  - If we can discover the structure of the IRF (deconvolution), we can convolve predictors with it to obtain a model of the response that takes diffusion directly into account

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Signal processing provides a framework for capturing temporal diffusion
  - Stimuli and responses can be recast as convolutionally-related signals
  - Relation described by an impulse response function (IRF)
  - If we can discover the structure of the IRF (deconvolution), we can convolve predictors with it to obtain a model of the response that takes diffusion directly into account

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Deconvolution is hard for psycholinguistic time series
- Major frameworks are discrete time
  - Finite impulse response models (FIR) (Dayal and MacGregor 1996)
  - Vector autoregression (VAR) (Sims 1980)
- Why is this a problem? **Variably-spaced events**

Motivation

- Deconvolution is hard for psycholinguistic time series
- Major frameworks are discrete time
  - Finite impulse response models (FIR) (Dayal and MacGregor 1996)
  - Vector autoregression (VAR) (Sims 1980)
- Why is this a problem? **Variably-spaced events**

Motivation

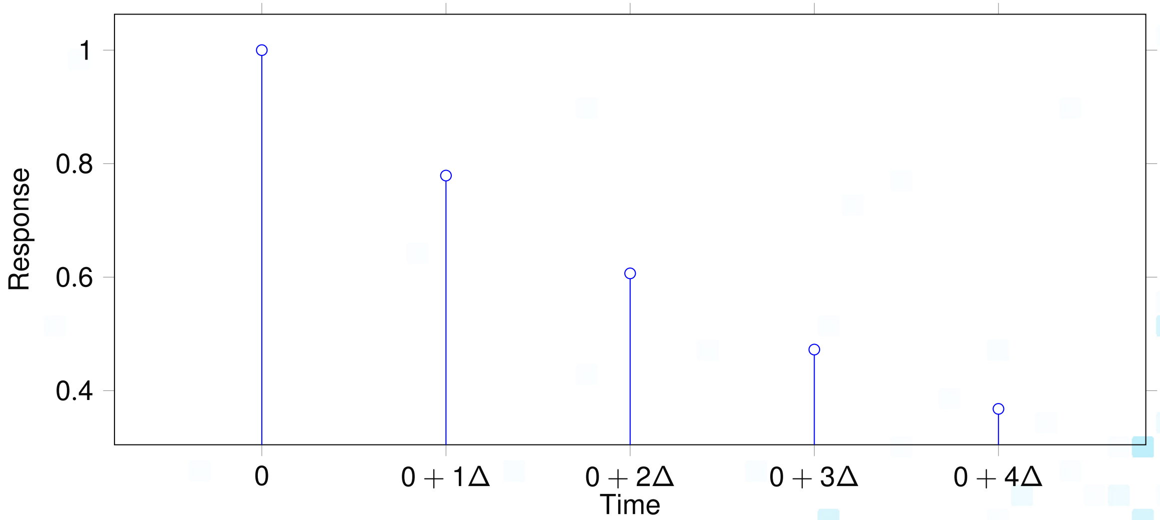
- Deconvolution is hard for psycholinguistic time series
- Major frameworks are discrete time
  - Finite impulse response models (FIR) (Dayal and MacGregor 1996)
  - Vector autoregression (VAR) (Sims 1980)
- Why is this a problem? **Variably-spaced events**

Motivation

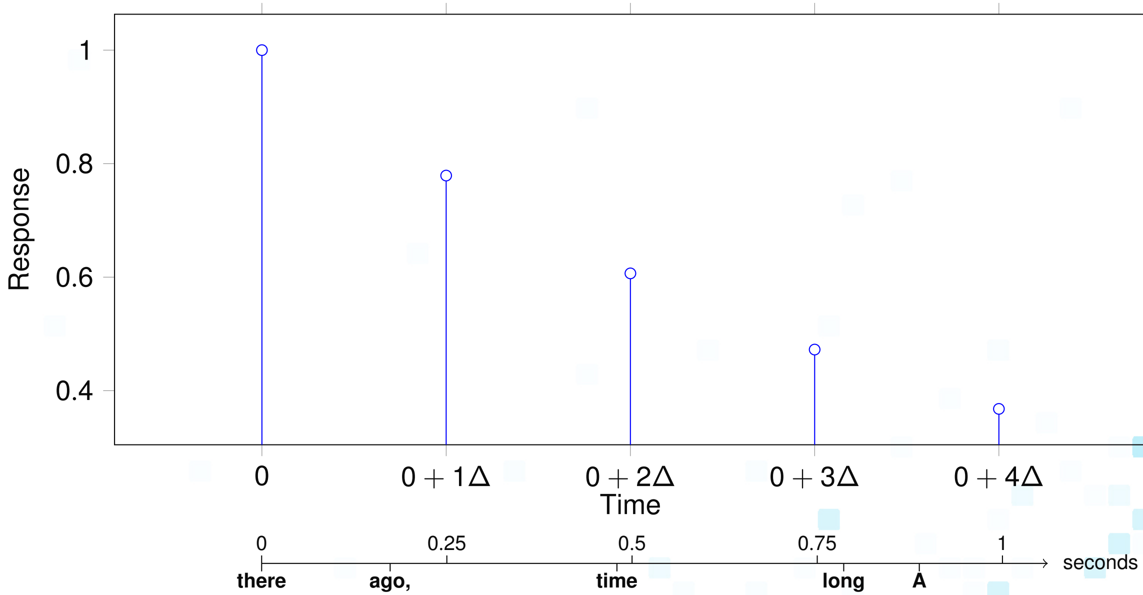
- Deconvolution is hard for psycholinguistic time series
- Major frameworks are discrete time
  - Finite impulse response models (FIR) (Dayal and MacGregor 1996)
  - Vector autoregression (VAR) (Sims 1980)
- Why is this a problem? **Variably-spaced events**

Motivation

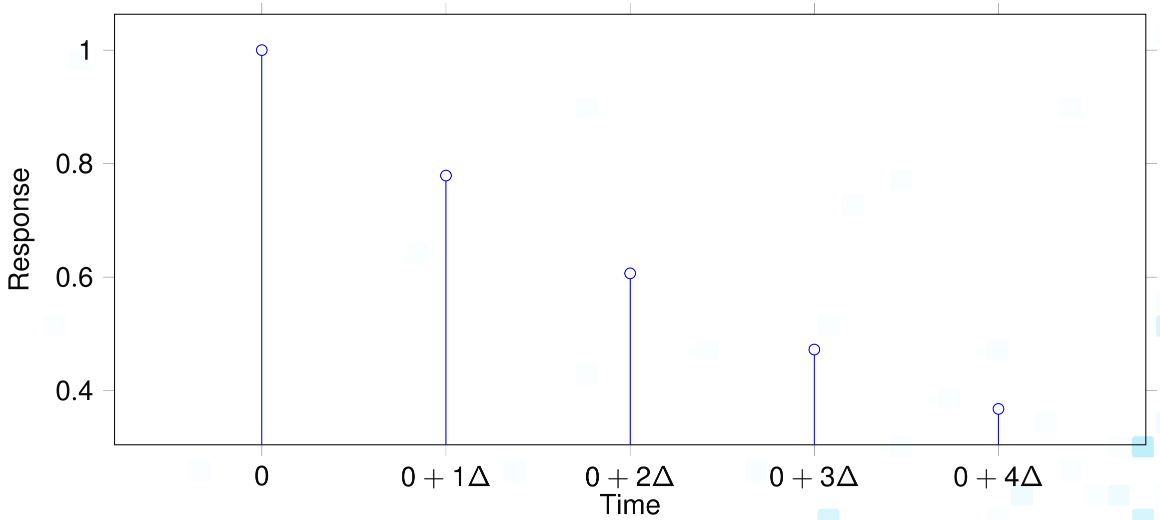
- Deconvolution is hard for psycholinguistic time series
- Major frameworks are discrete time
  - Finite impulse response models (FIR) (Dayal and MacGregor 1996)
  - Vector autoregression (VAR) (Sims 1980)
- Why is this a problem? **Variably-spaced events**

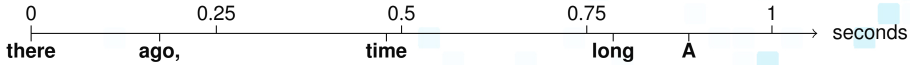


Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



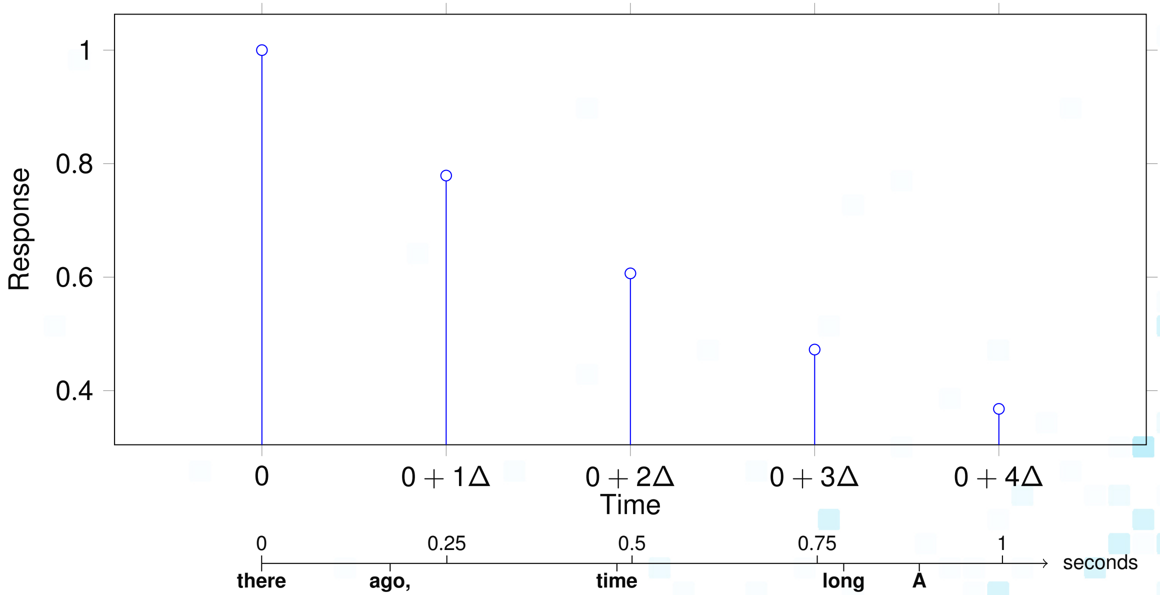
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,





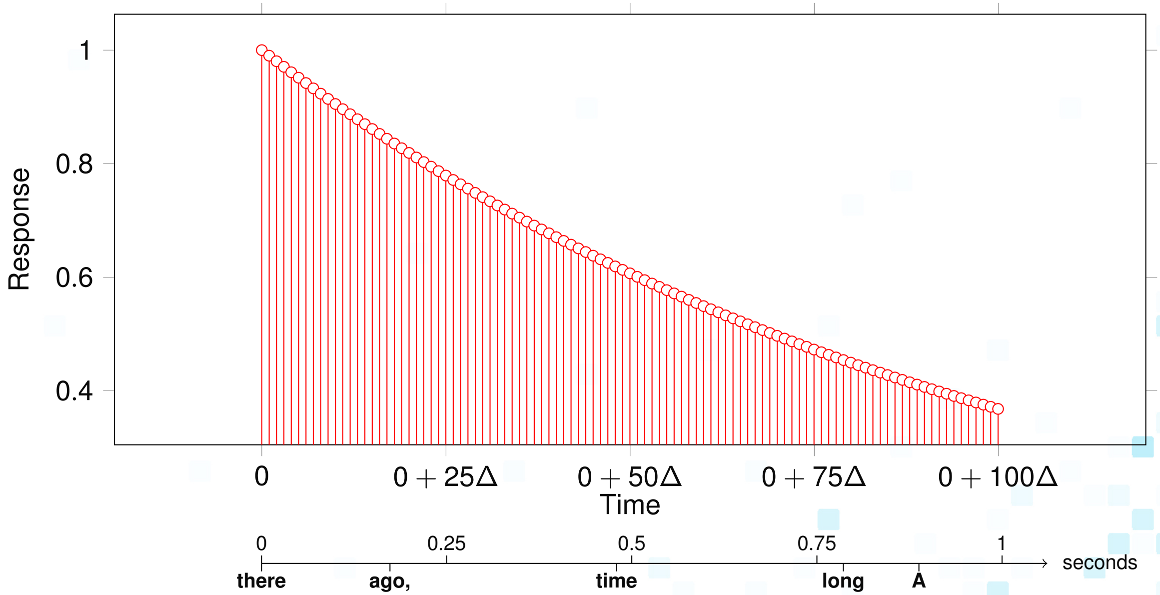
Variable spacing, Δ not fixed, can't deconvolve

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



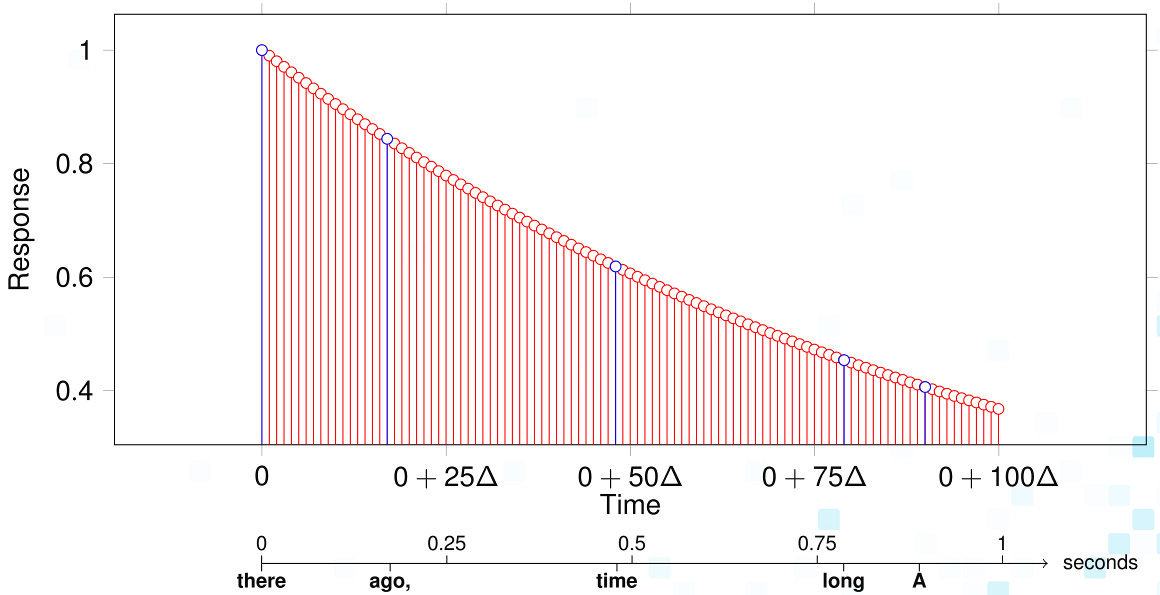
Sparse solution: Add *lots* of coefficients.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



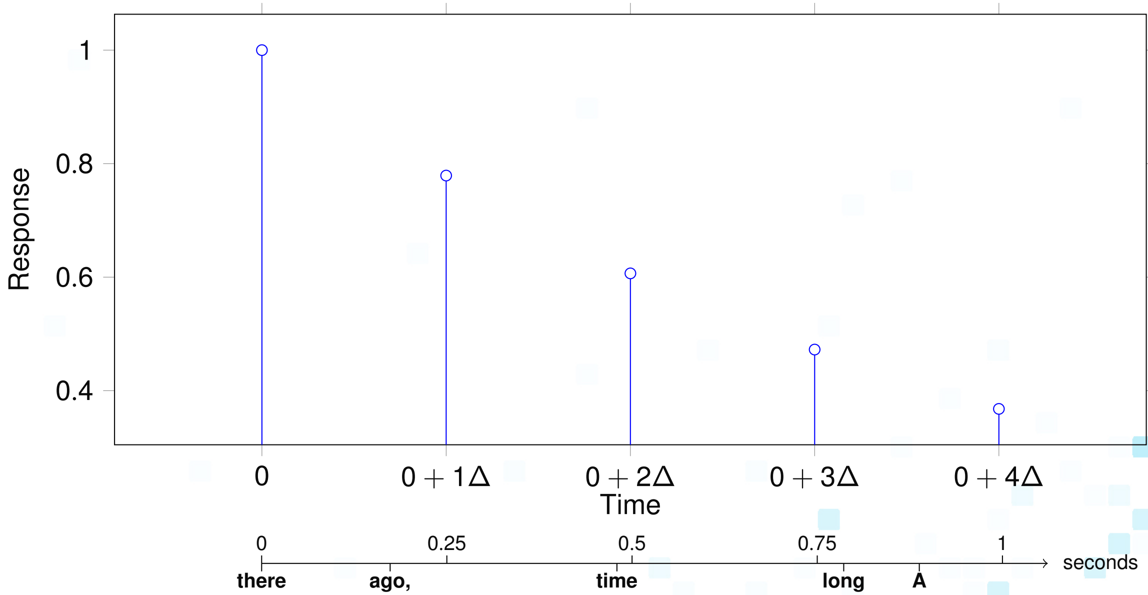
Sparse solution: Add *lots* of coefficients.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



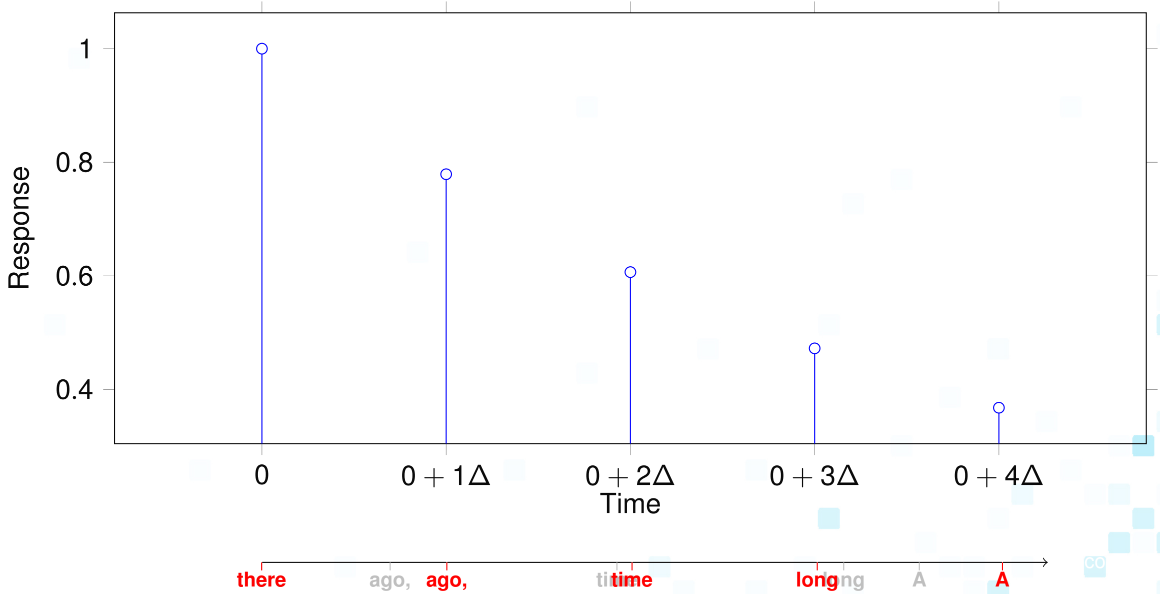
Sparse solution: Add *lots* of coefficients. $\Delta = 0.01$, but few coefficients have data.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



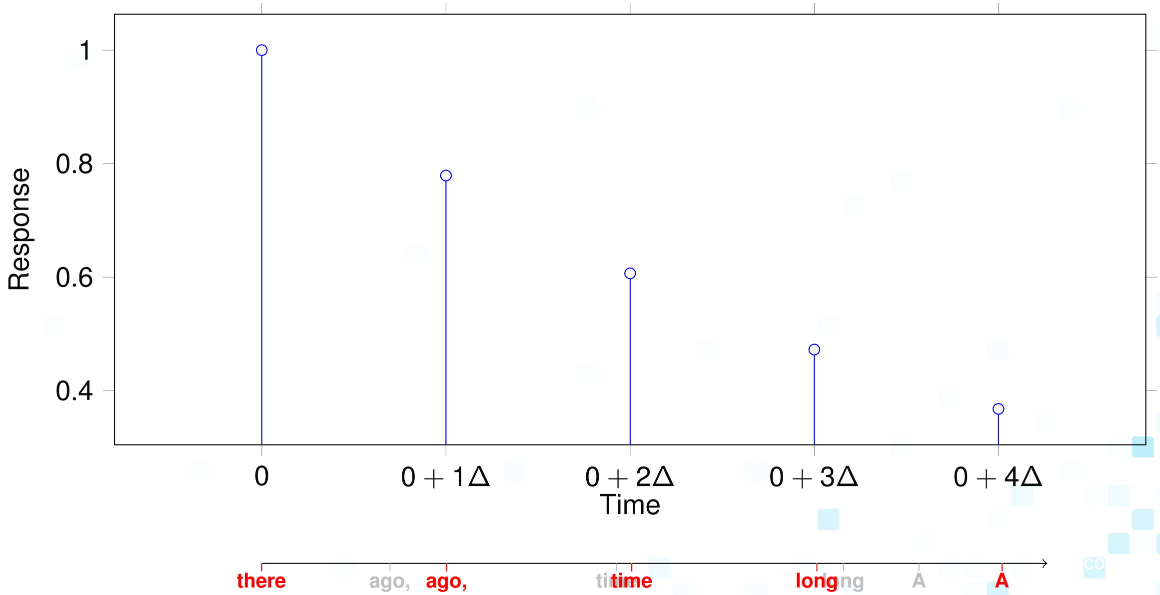
Distortionary solution: Delete temporal variation.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



Distortionary solution: Delete temporal variation.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



Distortionary solution: Delete temporal variation. Δ uninterpretable, time model broken.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- "Distortionary solution" might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicollinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Motivation

- “Distortionary solution” might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicollinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Motivation

- “Distortionary solution” might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicolinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Motivation

- "Distortionary solution" might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicolinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- "Distortionary solution" might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicolinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Motivation

- "Distortionary solution" might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicolinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- “Distortionary solution” might look familiar
- *Spillover* models like this are widely used in psycholinguistics (Erlich and Rayner 1983)
- Problems with spillover
  - Ignores temporal localization of events, only retains relative order
  - May introduce multicolinearity
  - Difficult to motivate choice of spillover configuration
  - Prone to overfitting and non-convergence, especially with random effects

Motivation

- Deconvolution is hard for psycholinguistic time series
- Failure to control for temporal diffusion can lead to misleading models

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Deconvolution is hard for psycholinguistic time series
- Failure to control for temporal diffusion can lead to misleading models

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

CASE IN POINT

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Shain et al. (2016): analysis of large SPR corpus (Futrell et al. 2018)
- Significant effects of **constituent wrap-up** and **dependency locality**
- First strong evidence of memory effects in broad-coverage sentence processing
- Paper has a couple of citations
- Accepted as a long-form talk at CUNY 2017

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Shain et al. (2016): analysis of large SPR corpus (Futrell et al. 2018)
- Significant effects of **constituent wrap-up** and **dependency locality**
- First strong evidence of memory effects in broad-coverage sentence processing
- Paper has a couple of citations
- Accepted as a long-form talk at CUNY 2017

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Shain et al. (2016): analysis of large SPR corpus (Futrell et al. 2018)
- Significant effects of **constituent wrap-up** and **dependency locality**
- First strong evidence of memory effects in broad-coverage sentence processing
- Paper has a couple of citations
- Accepted as a long-form talk at CUNY 2017

Motivation

- Shain et al. (2016): analysis of large SPR corpus (Futrell et al. 2018)
- Significant effects of **constituent wrap-up** and **dependency locality**
- First strong evidence of memory effects in broad-coverage sentence processing
- Paper has a couple of citations
- Accepted as a long-form talk at CUNY 2017

Motivation

- Shain et al. (2016): analysis of large SPR corpus (Futrell et al. 2018)
- Significant effects of **constituent wrap-up** and **dependency locality**
- First strong evidence of memory effects in broad-coverage sentence processing
- Paper has a couple of citations
- Accepted as a long-form talk at CUNY 2017

Motivation

	β-ms	t-value	p-value
Constituent wrap-up	1.54	8.15	**2.33e-14**
Dependency locality	1.10	6.48	**4.87e-10**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

	β-ms	t-value	p-value
Constituent wrap-up	1.54	8.15	**2.33e-14**
Dependency locality	1.10	6.48	**4.87e-10**

But after spilling over one baseline variable...

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

	β-ms	t-value	p-value
Constituent wrap-up	1.54	8.15	**2.33e-14**
Dependency locality	1.10	6.48	**4.87e-10**

But after spilling over one baseline variable...

~~Constituent wrap-up~~: $p = 0.816$

~~Dependency locality~~: $p = 0.370$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

	β-ms	t-value	p-value
Constituent wrap-up	1.54	8.15	**2.33e-14**
Dependency locality	1.10	6.48	**4.87e-10**

But after spilling over one baseline variable...

~~Constituent wrap-up~~: $p = 0.816$
~~Dependency locality~~: $p = 0.370$

Tiny tweak to timecourse modeling → huge impact on hypothesis testing

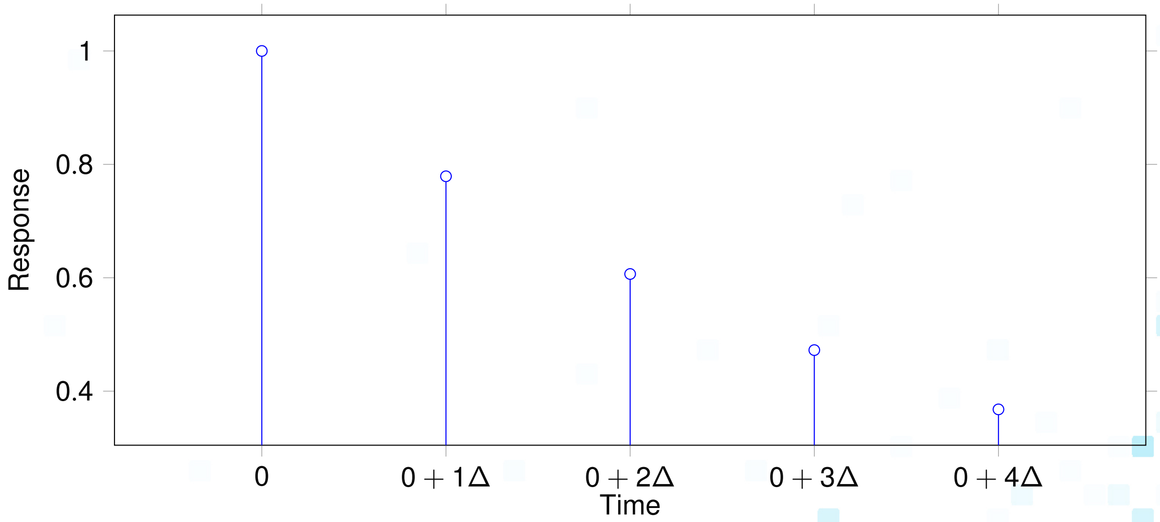
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

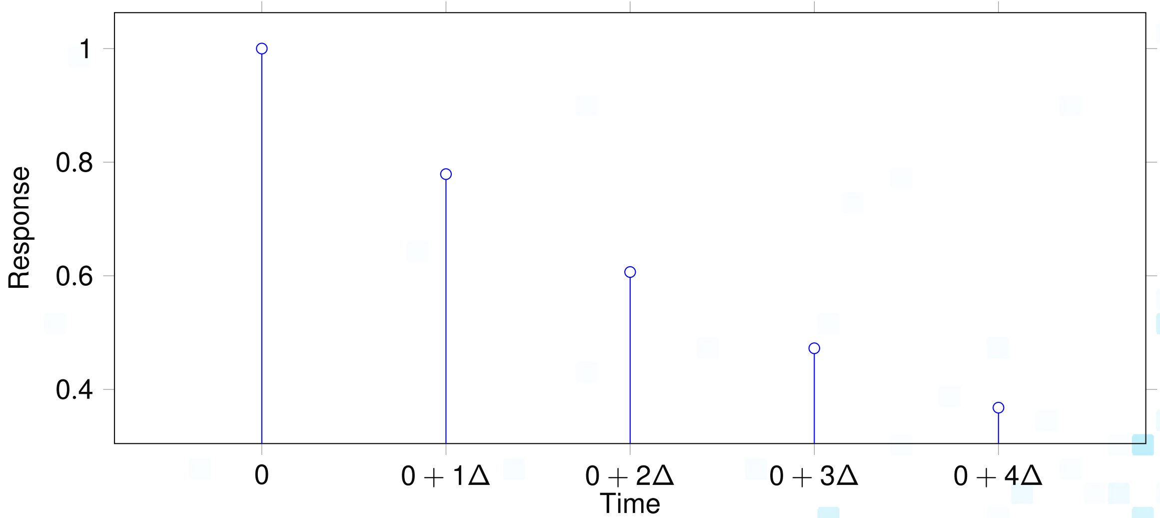
Deconvolution of psycholinguistic timecourses is both difficult and important.

What should we do?

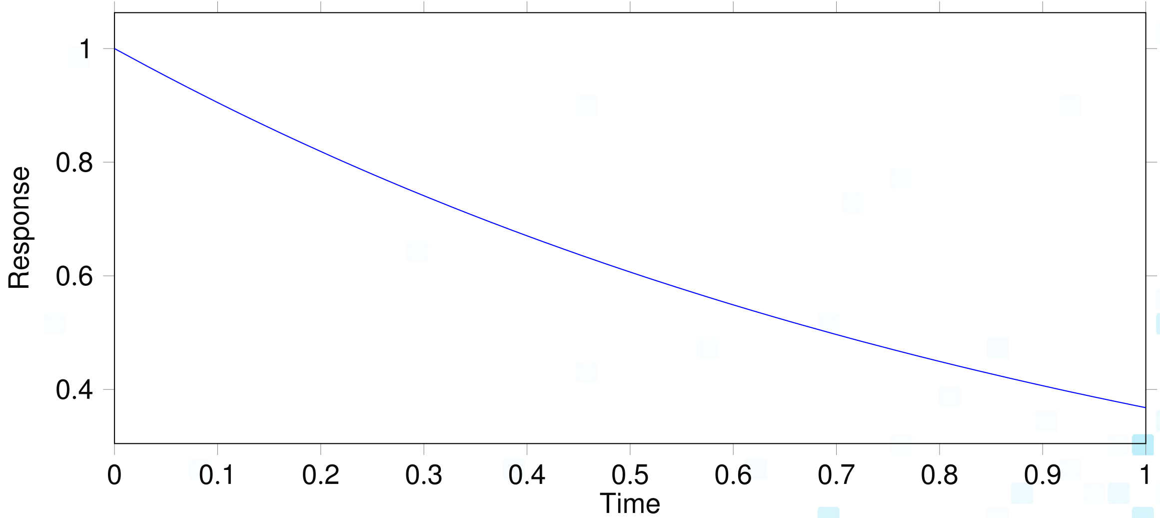
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



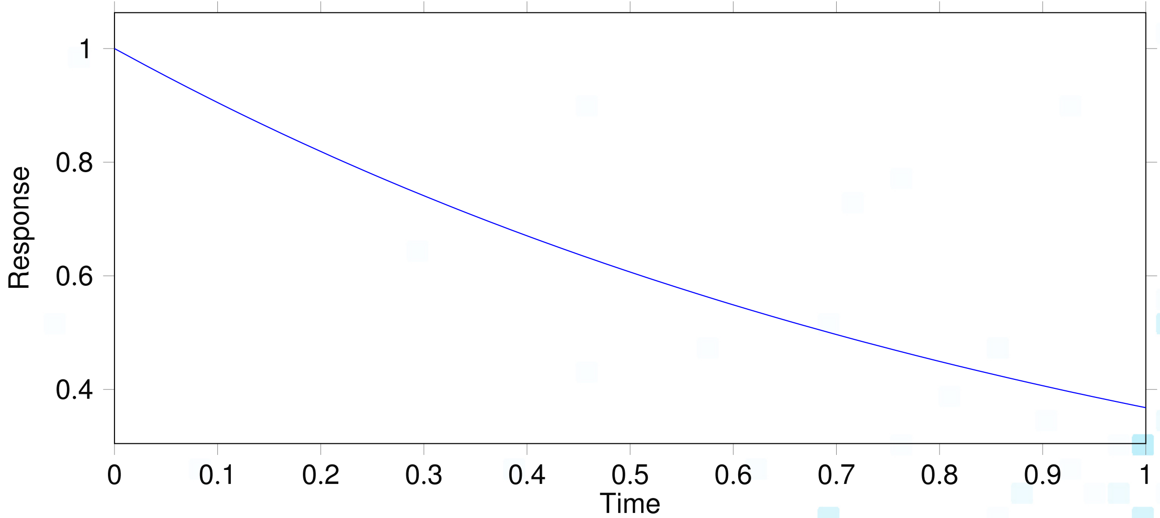
What if we had a continuous IRF?



What if we had a continuous IRF?

E.g. $f(x;\beta) = \beta e^{-\beta x}$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



What if we had a continuous IRF?

E.g. $f(x;\beta) = \beta e^{-\beta x}$

If we could fit β, we could predict the response *anywhere*

Motivation

- Continuous-time deconvolution would
  - Avoid discretizing time into lags
  - Support **variably-spaced** events
  - Support **unsynchronized** events
  - Apply without sparsity/distortion to any psycholinguistic time series

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Continuous-time deconvolution would
  - Avoid discretizing time into lags
  - Support **variably-spaced** events
  - Support **unsynchronized** events
  - Apply without sparsity/distortion to any psycholinguistic time series

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Continuous-time deconvolution would
  - Avoid discretizing time into lags
  - Support **variably-spaced** events
  - Support **unsynchronized** events
  - Apply without sparsity/distortion to any psycholinguistic time series

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Continuous-time deconvolution would
  - Avoid discretizing time into lags
  - Support **variably-spaced** events
  - Support **unsynchronized** events
  - Apply without sparsity/distortion to any psycholinguistic time series

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Continuous-time deconvolution would
  - Avoid discretizing time into lags
  - Support **variably-spaced** events
  - Support **unsynchronized** events
  - Apply without sparsity/distortion to any psycholinguistic time series

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1^{st} and 2^{nd} derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1st and 2nd derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1^{st} and 2^{nd} derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1st and 2nd derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1^{st} and 2^{nd} derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1st and 2nd derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Motivation

- Until recently, continuous-time deconvolution was hard because non-linear in its parameters
- Estimators would have to be derived by hand
  - Derive likelihood function (depends on IRF)
  - Find its 1^{st} and 2^{nd} derivatives w.r.t. all parameters
  - Use derivatives to compute maximum likelihood estimators
  - Repeat for new model
- Recent developments in machine learning allow us to avoid this through **autodifferentiation** and **stochastic optimization**

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Proposal: Deconvolutional Time Series Regression

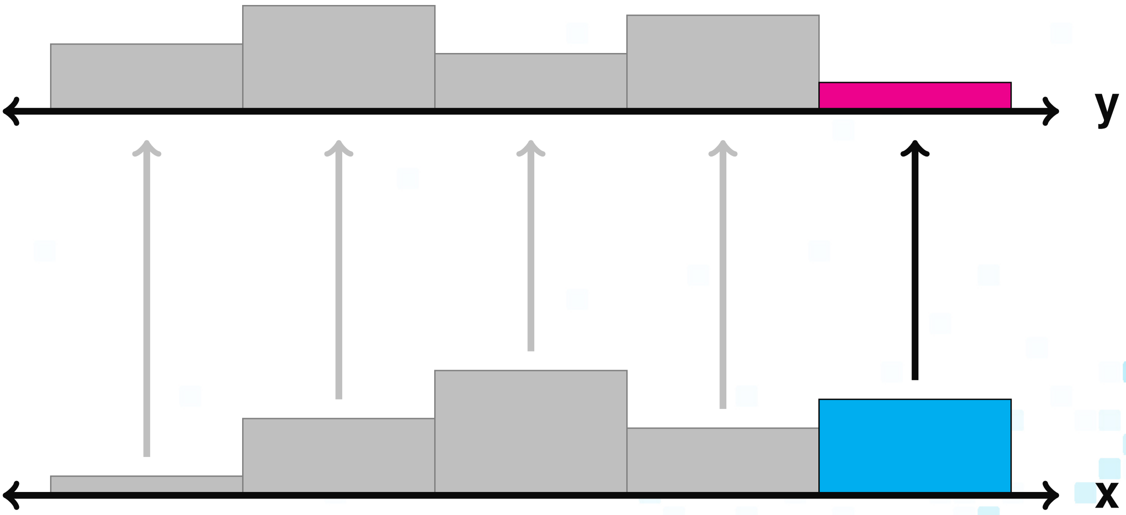
- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $O(1)$ model complexity on num. timesteps

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

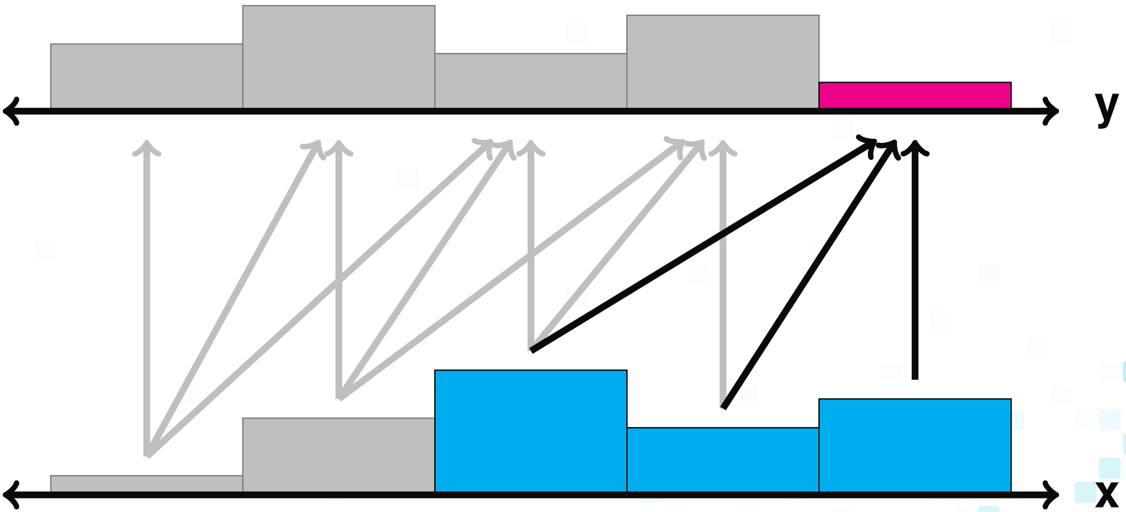
Proposal: Deconvolutional Time Series Regression

- Jointly fits:
  - Continuous-time parametric IRFs for each predictor
  - Linear model on convolved predictors
- Uses autodifferentiation and gradient-based
- Applies to any time series using any set of parametric IRF kernels optimization
- Provides an interpretable model that directly estimates temporal diffusion
- $\mathcal{O}(1)$ model complexity on num. timesteps

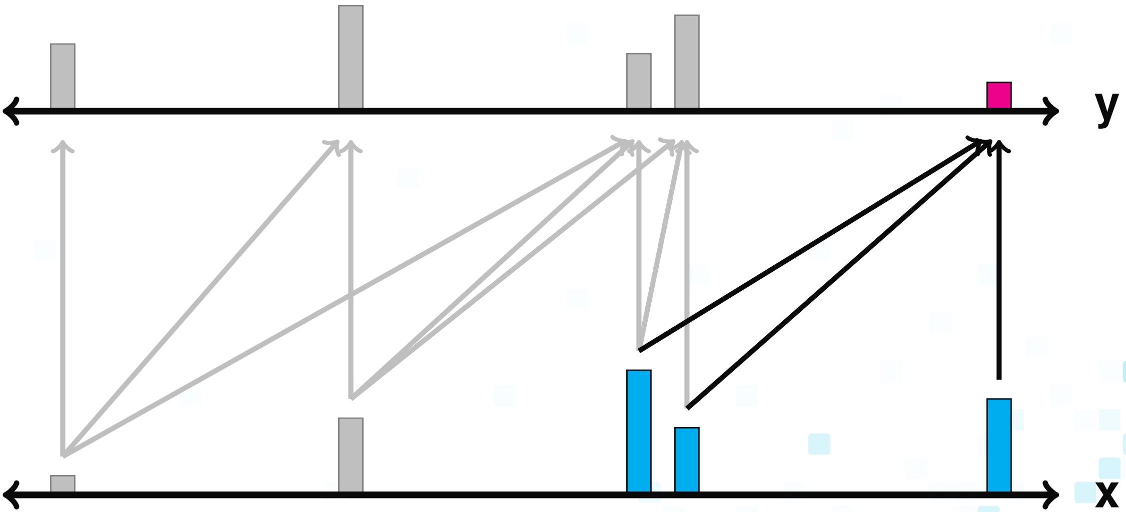
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



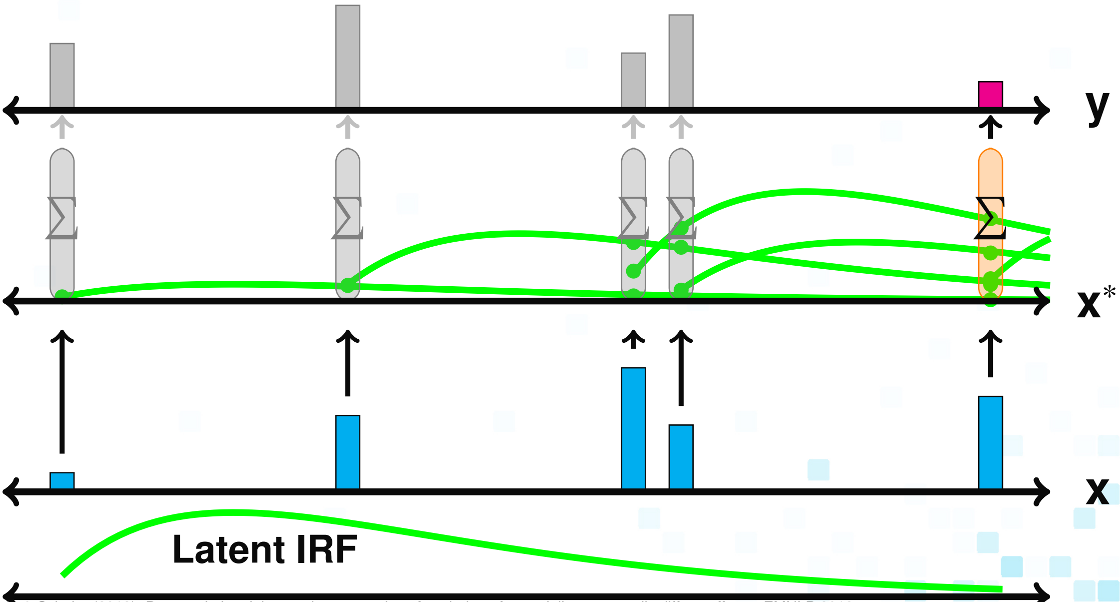
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,



Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,



Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Proposal: Deconvolutional Time Series Regression

- Expands range of application of deconvolutional modeling (e.g. to reading)
- Provides high-resolution estimates of temporal dynamics
- Documented open-source Python package supports
  - Mixed effects modeling (intercepts, slopes, and IRF parameters)
  - Various IRF kernels (and more coming)
  - Non-parametric IRFs through spline kernels
  - Composition of IRF kernels
  - MLE, Bayesian, and variational Bayesian inference modes
- `https://github.com/coryshain/dtsr`

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

DTSR Implementation Used Here

- ShiftedGamma IRF kernel

$$f(x;\alpha,\beta,\delta) = \frac{\beta^{\alpha}(x-\delta)^{\alpha-1}e^{-\beta(x-\delta)}}{\Gamma(\alpha)}$$

- Black box variational inference (BBVI)
- Implemented in Tensorflow (Abadi et al. 2015) and Edward (Tran et al. 2016)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

DTSR Implementation Used Here

- ShiftedGamma IRF kernel

$$f(x; \alpha, \beta, \delta) = \frac{\beta^{\alpha}(x-\delta)^{\alpha-1}e^{-\beta(x-\delta)}}{\Gamma(\alpha)}$$

- Black box variational inference (BBVI)
- Implemented in Tensorflow (Abadi et al. 2015) and Edward (Tran et al. 2016)

DTSR Implementation Used Here

- ShiftedGamma IRF kernel

$$f(x;\alpha,\beta,\delta) = \frac{\beta^{\alpha}(x-\delta)^{\alpha-1}e^{-\beta(x-\delta)}}{\Gamma(\alpha)}$$

- Black box variational inference (BBVI)
- Implemented in Tensorflow (Abadi et al. 2015) and Edward (Tran et al. 2016)

Synthetic Evaluation

- Sanity check: Can DTSR recover known IRFs?
- Generate data from a model with known convolutional structure
- Fit DTSR to that data and compare estimates to ground truth

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Synthetic Evaluation

- Sanity check: Can DTSR recover known IRFs?
- Generate data from a model with known convolutional structure
- Fit DTSR to that data and compare estimates to ground truth

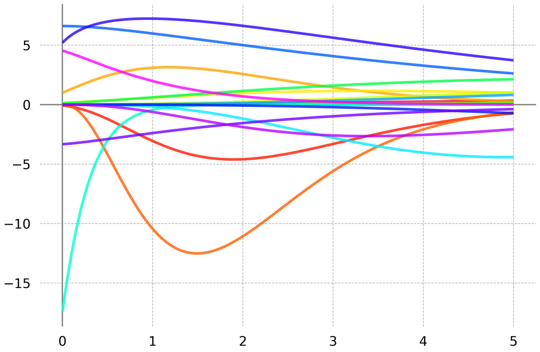
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Synthetic Evaluation

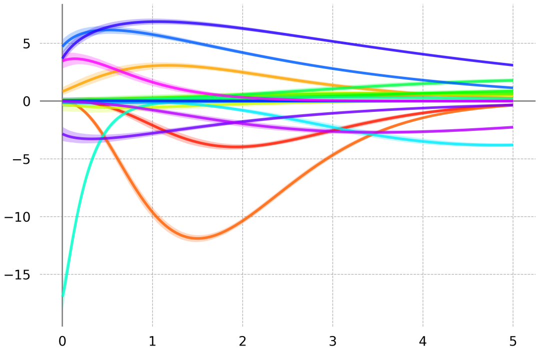
- Sanity check: Can DTSR recover known IRFs?
- Generate data from a model with known convolutional structure
- Fit DTSR to that data and compare estimates to ground truth

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Synthetic Evaluation



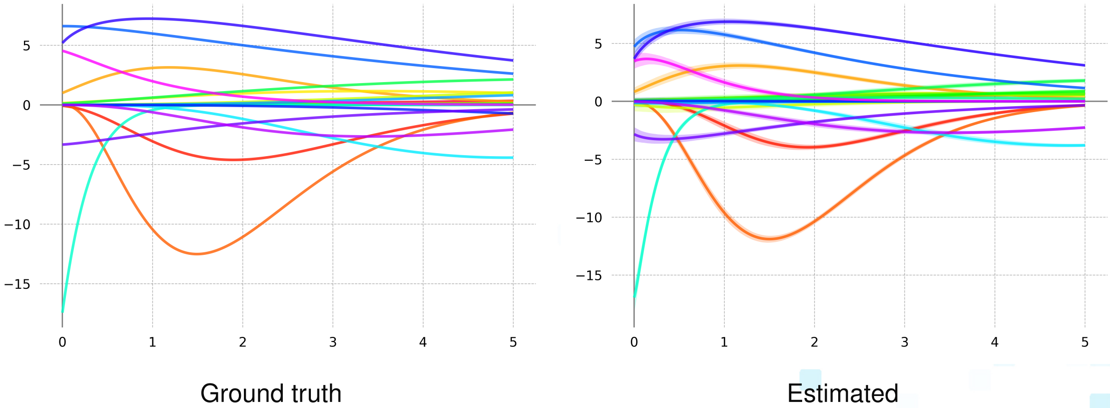
Ground truth



Estimated

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

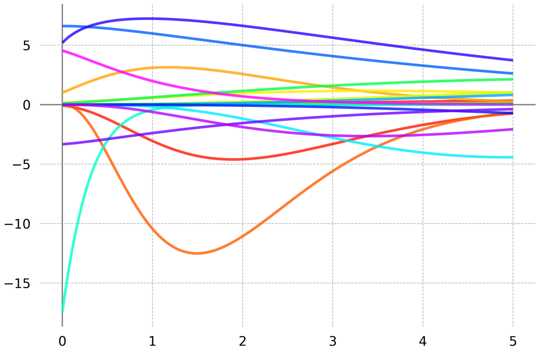
Synthetic Evaluation



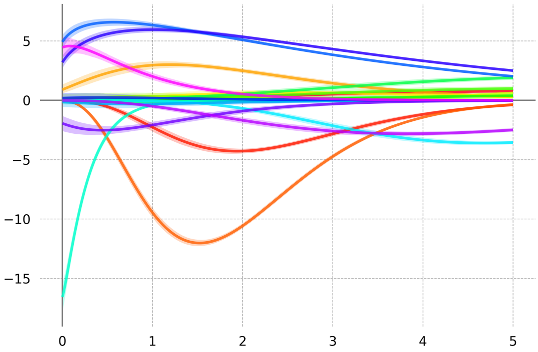
$$\rho = 0$$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Synthetic Evaluation



Ground truth

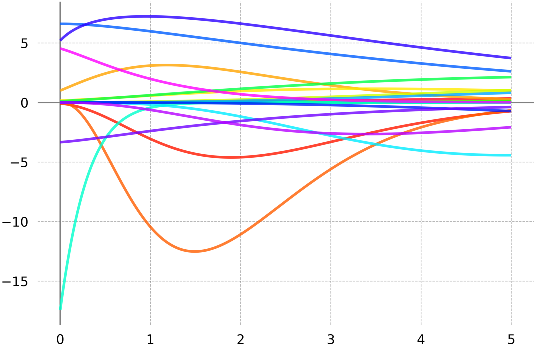


Estimated

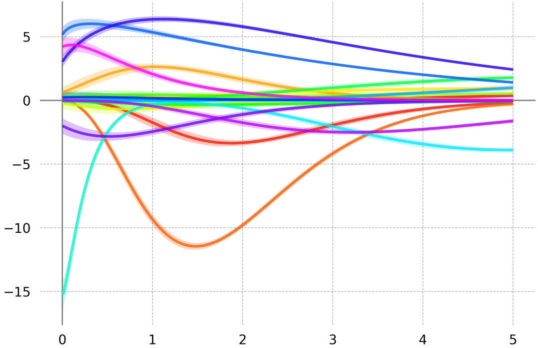
$$\rho = 0.25$$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Synthetic Evaluation



Ground truth

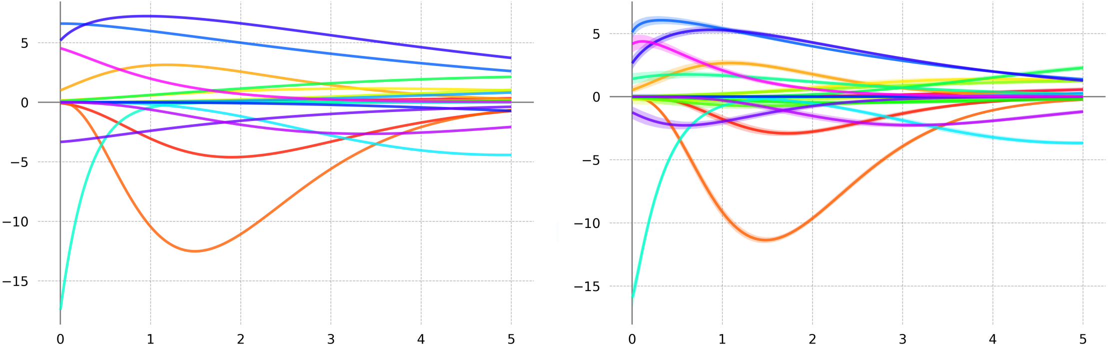


Estimated

$\rho = 0.5$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Synthetic Evaluation



Ground truth

Estimated

$$\rho = 0.75$$

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Synthetic Evaluation

- DTSR can recover known IRFs with high fidelity
- Estimates are robust to multicolinearity

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Synthetic Evaluation

- DTSR can recover known IRFs with high fidelity
- Estimates are robust to multicolinearity

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Datasets:
  - Natural Stories (SPR) (Futrell et al. 2018)
  - Dundee (ET) (Kennedy, Pynte, and Hill 2003)
  - UCL (ET) (Frank et al. 2013)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Reading Times

- Convolved predictors
  - Saccade length (eye-tracking only)
  - Word length
  - Unigram logprob
  - 5-gram surprisal
  - Rate (DTSR only)
- Linear predictors
  - Sentence position
  - Trial
- Response: Log reading times (go-past for eye-tracking)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Reading Times

- More on Rate:
  - Rate predictor is an intercept (vector of 1's) that gets convolved with an IRF
  - Captures effects of stimulus *timing* independently of stimulus *properties*
  - Only detectable through deconvolution

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- More on Rate:
  - Rate predictor is an intercept (vector of 1's) that gets convolved with an IRF
  - Captures effects of stimulus *timing* independently of stimulus *properties*
  - Only detectable through deconvolution

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- More on Rate:
  - Rate predictor is an intercept (vector of 1's) that gets convolved with an IRF
  - Captures effects of stimulus *timing* independently of stimulus *properties*
  - Only detectable through deconvolution

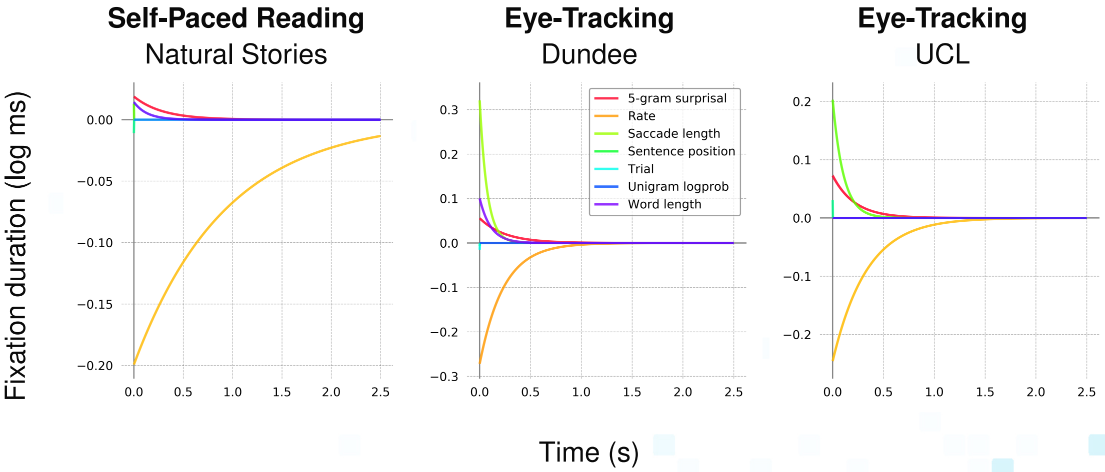
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Reading Times

- More on Rate:
  - Rate predictor is an intercept (vector of 1's) that gets convolved with an IRF
  - Captures effects of stimulus *timing* independently of stimulus *properties*
  - Only detectable through deconvolution

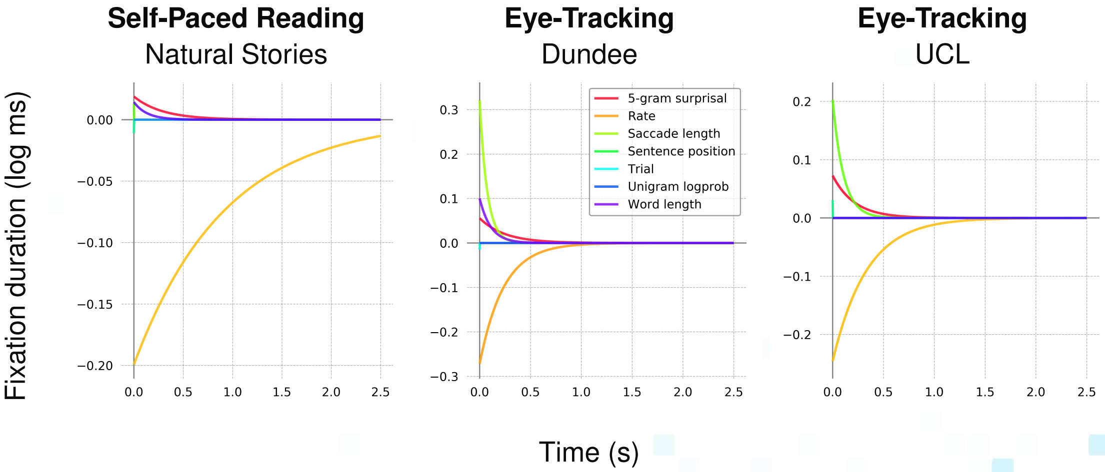
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Fitted IRF



Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

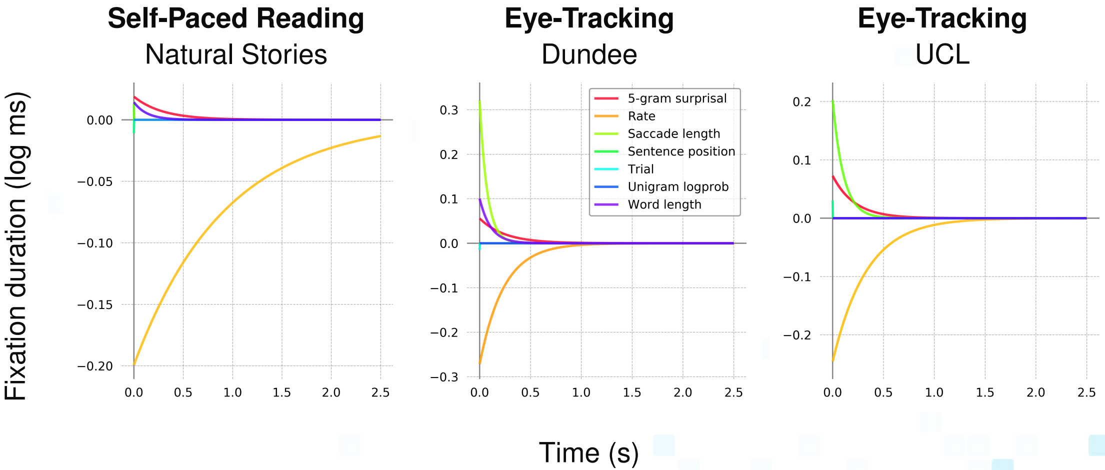
Naturalistic Evaluation: Fitted IRF



Large negative influence of Rate (convolved intercept) suggests inertia

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

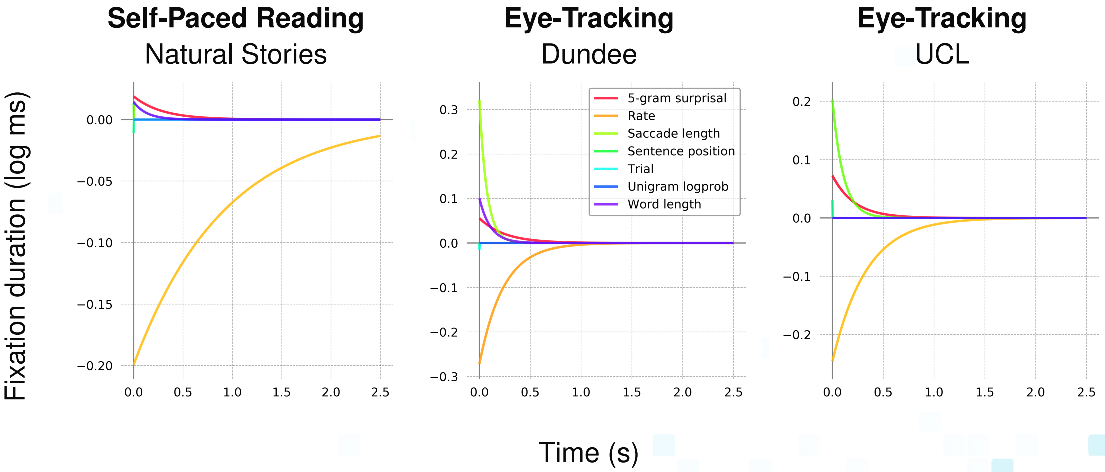
Naturalistic Evaluation: Fitted IRF



Diffusion mostly restricted to first second after stimulus presentation

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

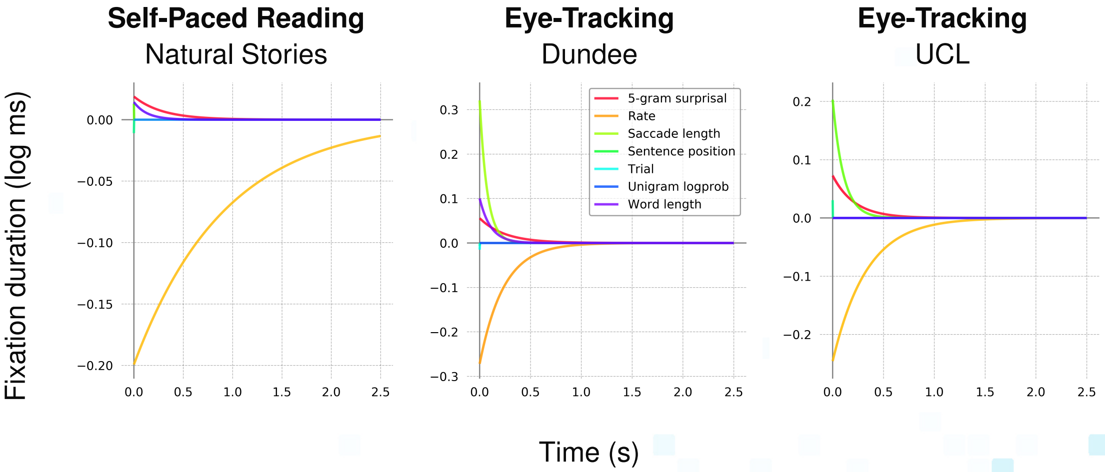
Naturalistic Evaluation: Fitted IRF



Top-down response slower than bottom-up (surp vs. word/sac. len) (Friederici 2002)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

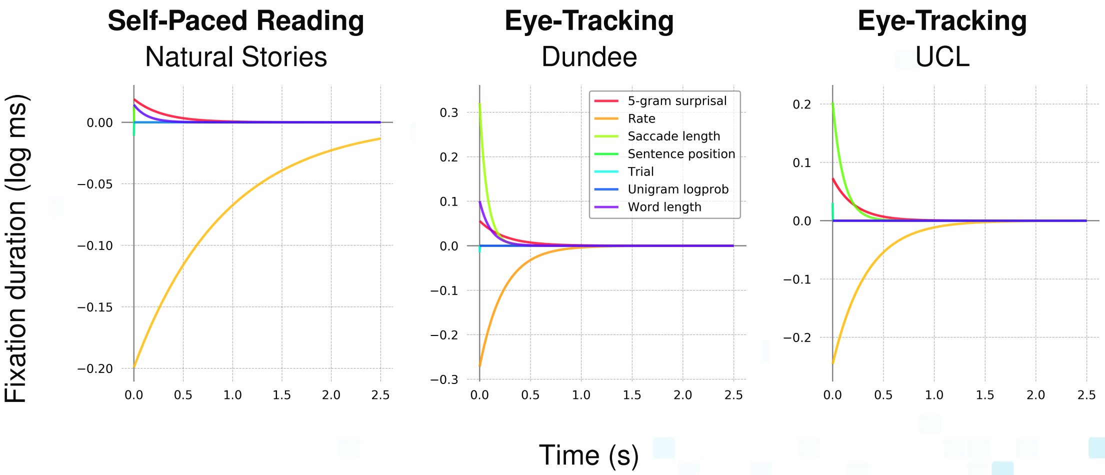
Naturalistic Evaluation: Fitted IRF



Similar temporal profile across eye-tracking corpora

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Fitted IRF

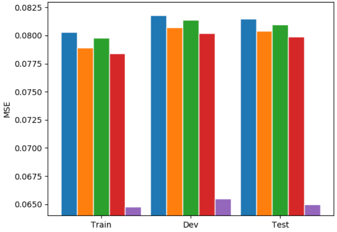


Null influence of unigram logprob (c.f. e.g. Levy 2008; Staub 2015)

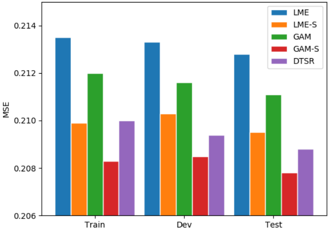
Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: System Comparison

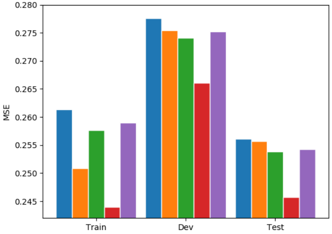
Natural Stories



Dundee



UCL



Mean squared prediction error (MSPE), DTSR vs. competitors
LME (blue); LME-S (orange); GAM (green); GAM-S (red); DTSR (purple)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Naturalistic Evaluation: Summary

- Estimated IRFs shed new light on temporal dynamics in naturalistic reading
- Estimates are plausible, replicable, and fine-grained
- Models show high quality prediction performance, validating IRFs

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Summary

- Estimated IRFs shed new light on temporal dynamics in naturalistic reading
- Estimates are plausible, replicable, and fine-grained
- Models show high quality prediction performance, validating IRFs

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Naturalistic Evaluation: Summary

- Estimated IRFs shed new light on temporal dynamics in naturalistic reading
- Estimates are plausible, replicable, and fine-grained
- Models show high quality prediction performance, validating IRFs

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Hypothesis Testing

So how do I test a claim using DTSR?

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- DTSR stochastically optimizes over a non-convex likelihood surface
  - Nearly ubiquitous property of modern machine learning algorithms
  - Introduces possibility of *estimation noise*
    - Convergence to a non-global optimum
    - Imperfect convergence to an optimum
    - Evaluation using Monte Carlo sampling (Bayesian only)
- Estimates and training predictions/likelihoods are not guaranteed to be globally optimal
- Differences between models may be influenced artifacts of fitting procedure

Hypothesis Testing

- Despite not being provably optimal
  - Synthetic results suggest DTSR does recover model near-optimally
  - We want to understand a non-linear and non-convex world

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Despite not being provably optimal
  - Synthetic results suggest DTSR does recover model near-optimally
  - We want to understand a non-linear and non-convex world

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Despite not being provably optimal
  - Synthetic results suggest DTSR does recover model near-optimally
  - We want to understand a non-linear and non-convex world

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
    - **Spirit:** Machine learning “bakeoff”
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
    - **Spirit:** Machine learning “bakeoff”
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
    - **Spirit:** Pre-convolution with canonical HRF in fMRI
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing

- Three frameworks for using DTSR in hypothesis tests
  - Directly compare DTSR models (permutation test)
    - **Spirit:** Machine learning “bakeoff”
  - Use DTSR to *transform* predictors as inputs to linear models (2-step test)
    - **Spirit:** Pre-convolution with canonical HRF in fMRI
  - Use DTSR to (1) compute a Rate predictor and (2) motivate spillover structure
    - **Spirit:** Exploratory data analysis

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Hypothesis Testing: Example

Test	*p*-value
Permutation	9.99e-05***
2-Step	**TBD**

In-sample test for effect of *Surprisal* in Natural Stories

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Other applications of DTSR

- **Other response measures:** E.g. HRF deconvolution with naturalistic stimuli
- **2D predictors:** E.g. effects of word cosine similarities
- **Composed IRFs**: E.g. separating neural and hemodynamic responses in fMRI
- **Spline kernels**: E.g. response shape discovery
- ...

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Other applications of DTSR

- **Other response measures:** E.g. HRF deconvolution with naturalistic stimuli
- **2D predictors:** E.g. effects of word cosine similarities
- **Composed IRFs**: E.g. separating neural and hemodynamic responses in fMRI
- **Spline kernels**: E.g. response shape discovery
- ...

Other applications of DTSR

- **Other response measures:** E.g. HRF deconvolution with naturalistic stimuli
- **2D predictors:** E.g. effects of word cosine similarities
- **Composed IRFs**: E.g. separating neural and hemodynamic responses in fMRI
- **Spline kernels**: E.g. response shape discovery
- ...

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Other applications of DTSR

- **Other response measures:** E.g. HRF deconvolution with naturalistic stimuli
- **2D predictors:** E.g. effects of word cosine similarities
- **Composed IRFs**: E.g. separating neural and hemodynamic responses in fMRI
- **Spline kernels**: E.g. response shape discovery
- ...

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Other applications of DTSR

- **Other response measures:** E.g. HRF deconvolution with naturalistic stimuli
- **2D predictors:** E.g. effects of word cosine similarities
- **Composed IRFs**: E.g. separating neural and hemodynamic responses in fMRI
- **Spline kernels**: E.g. response shape discovery
- ...

You too can DTSR!

Demo...

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Conclusion

- DTSR
  - Provides plausible, replicable, and high resolution estimates of temporal dynamics
  - Affords new insights into the temporal dynamics of reading behavior
  - Recovers known ground-truth IRFs with high fidelity
  - Applies to variably-spaced time series
  - Can help avoid spurious findings due to poor control of temporal diffusion
  - Can be integrated into various hypothesis testing frameworks
  - Is supported by documented open-source software

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,

Thank you!

Acknowledgments:

Reviewers and participants in *CUNY 2018* and *EMNLP 2018*.

This work was supported by National Science Foundation grants #1551313 and #1816891. All views expressed are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

- Abadi, Martín et al. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. URL: http://download.tensorflow.org/paper/whitepaper2015.pdf.
- Dayal, Bhupinder S and John F MacGregor (1996). "Identification of finite impulse response models: methods and robustness issues". In: Industrial & engineering chemistry research 35.11, pp. 4078–4090.
- Erlich, Kate and Keith Rayner (1983). "Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing". In: Journal of Verbal Learning & Verbal Behavior 22, pp. 75–87.
- Frank, Stefan L et al. (2013). "Reading time data for evaluating broad-coverage models of English sentence processing". In: Behavior Research Methods 45.4, pp. 1182–1190.
- Friederici, Angela D (2002). "Towards a neural basis of auditory sentence processing". In: Trends in Cognitive Sciences 6.2, pp. 78–84.

References

- Futrell, Richard et al. (2018). “The Natural Stories Corpus”. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation Ed. by Nicoletta Calzolari et al. Paris, France: European Language Resources Association (ELRA). ISBN: 979-10-95546-00-9.
- Kennedy, Alan, James Pynte, and Robin Hill (2003). “The Dundee corpus”. In: Proceedings of the 12th European conference on eye movement.
- Levy, Roger (2008). “Expectation-based syntactic comprehension”. In: Cognition 106.3, pp. 1126–1177.
- Shain, Cory et al. (2016). “Memory access during incremental sentence processing causes reading time latency”. In: Proceedings of the Computational Linguistics for Linguistic Complexity Workshop. Association for Computational Linguistics, pp. 49–58.
- Sims, Christopher A (1980). “Macroeconomics and reality”. In: Econometrica: Journal of the Econometric Society, pp. 1–48.

References

Staub, Adrian (2015). "The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation". In: Language and Linguistics Compass 9.8, pp. 311–327.

Tran, Dustin et al. (2016). "Edward: A library for probabilistic modeling, inference, and criticism". In: arXiv preprint arXiv:1610.09787.

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Appendix: Synthetic data generation procedure

- 10,000 data points 100ms apart
- 20 randomly sampled covariates $\sim \mathcal{N}(0, 1)$
- 20 unique coefficients $\mathcal{U}(-50, 50)$
- 20 unique IRF
  - $k \sim \mathcal{U}(1, 6)$
  - $\theta \sim \mathcal{U}(0, 5)$
  - $\delta \sim \mathcal{U}(0, 1)$
- Noise added $\sim \mathcal{N}(0, 20^2)$
- DTSR history window clipped at 128 observations

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Appendix: Reading time experiments

- **Natural Stories** (Futrell et al. 2018)
  - Constructed narratives, self-paced reading, 181 subjects, 485 sentences, 10,245 tokens, 848,768 fixation events
  - Post-processing: Removed sentence boundaries, events for which subjects missed 4+ comprehension questions and fixations $<$ 100 ms or $>$ 3000 ms.
- **Dundee** (Kennedy, Pynte, and Hill 2003)
  - Newspaper editorials, eye-tracking, 10 subjects, 2,368 sentences, 51,502 tokens, 260,065 fixation events
  - Post-processing: Removed document, screen, sentence, and line boundaries
- **UCL** (Frank et al. 2013)
  - Sentences from novels presented in isolation, eye-tracking, 42 subjects, 205 sentences, 1,931 tokens, 53,070 fixation events
  - Post-processing: Removed sentence boundaries

Appendix: Reading time experiments

- Baselines
  - LME (`lme4`) and GAM (`mgcv`)
  - By-subject intercepts and slopes
  - Spillover variants
    - No predictors spilled over
    - Spillover 0-3 for each predictor (-*S*)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018*.,

Appendix: Reading time experiments

- Data split
  - Train (50%)
  - Dev (25%)
  - Test (25%)

Shain & Schuler (2018). Deconvolutional time series regression: A technique for modeling temporally diffuse effects. *EMNLP 2018.*,