Lecture III: Review of Classic Quadratic Variation Results and Relevance to Statistical Inference in Finance

Christopher P. Calderon

Rice University / Numerica Corporation

Research Scientist

Outline

I Refresher on Some Unique Properties of Brownian Motion

II Stochastic Integration and Quadratic Variation of SDEs

III Demonstration of How Results Help Understand ”Realized Variation” Discretization Error and Idea Behind “Two Scale Realized Volatility Estimator”

Part I

Refresher on Some Unique Properties of Brownian Motion

Outline

For Items I & II Draw Heavily from:

Protter, P. (2004) *Stochastic Integration and Differential Equations*, Springer-Verlag, Berlin.
Steele, J.M. (2001) *Stochastic Calculus and Financial Applications*, Springer, New York.

For Item III Highlight Results from:

Barndorff-Nielsen, O. & Shepard, N. (2003) *Bernoulli* **9** 243-265.
Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

Assuming Some Basic Familiarity with Brownian Motion (B.M.)

- Stationary Independent Increments
- Increments are Gaussian $B_t - B_s \sim \mathcal{N}(0, t-s)$
- Paths are Continuous but "Rough" / "Jittery" (Not Classically Differentiable)
- Paths are of Infinite Variation so "Funny" Integrals Used in Stochastic Integration, e.g.

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t)$$

Assuming Some Basic Familiarity with Brownian Motion (B.M.)

The Material in Parts I and II are “Classic” Fundamental & Established Results but Set the Stage to Understand Basics in Part III.

The References Listed at the Beginning Provide a Detailed Mathematical Account of the Material I Summarize Briefly Here.

A Classic Result Worth Reflecting On

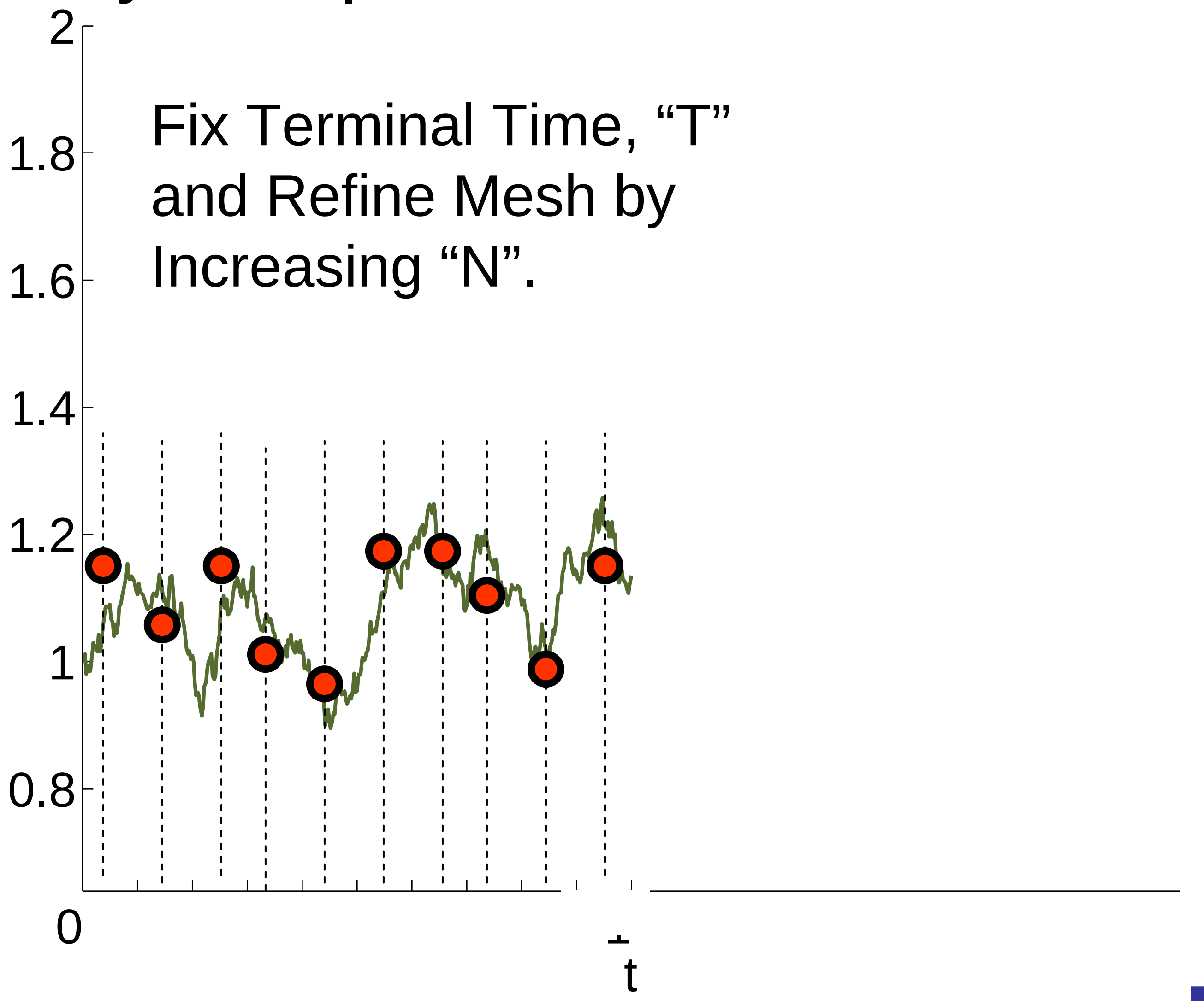
$$\lim_{N \to \infty} \sum_{i=1}^{N} (B_{t_i} - B_{t_{i-1}})^2 = T$$

$$t_i = i\frac{T}{N}$$

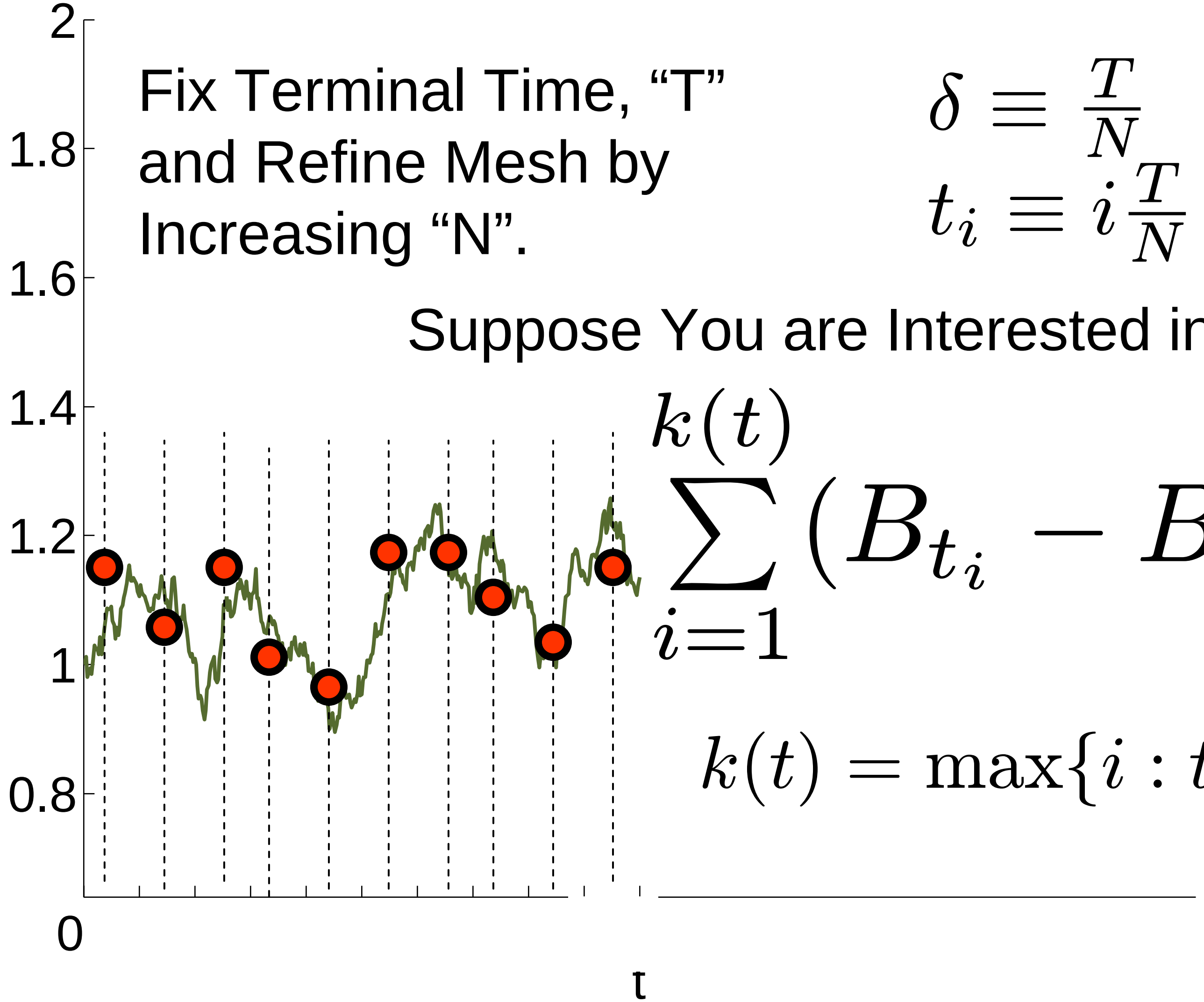
Implies Something Not Intuitive to Many Unfamiliar with Brownian Motion …

Discretely Sampled Brownian Motion Paths

Shift IC to Zero if "Standard Brownian Motion" Desired



Discretely Sampled Brownian Motion Paths



Fix Terminal Time, "T" and Refine Mesh by Increasing "N".

$$\delta \equiv \frac{T}{N}$$

$$t_i \equiv i\frac{T}{N}$$

Suppose You are Interested in Limit of:

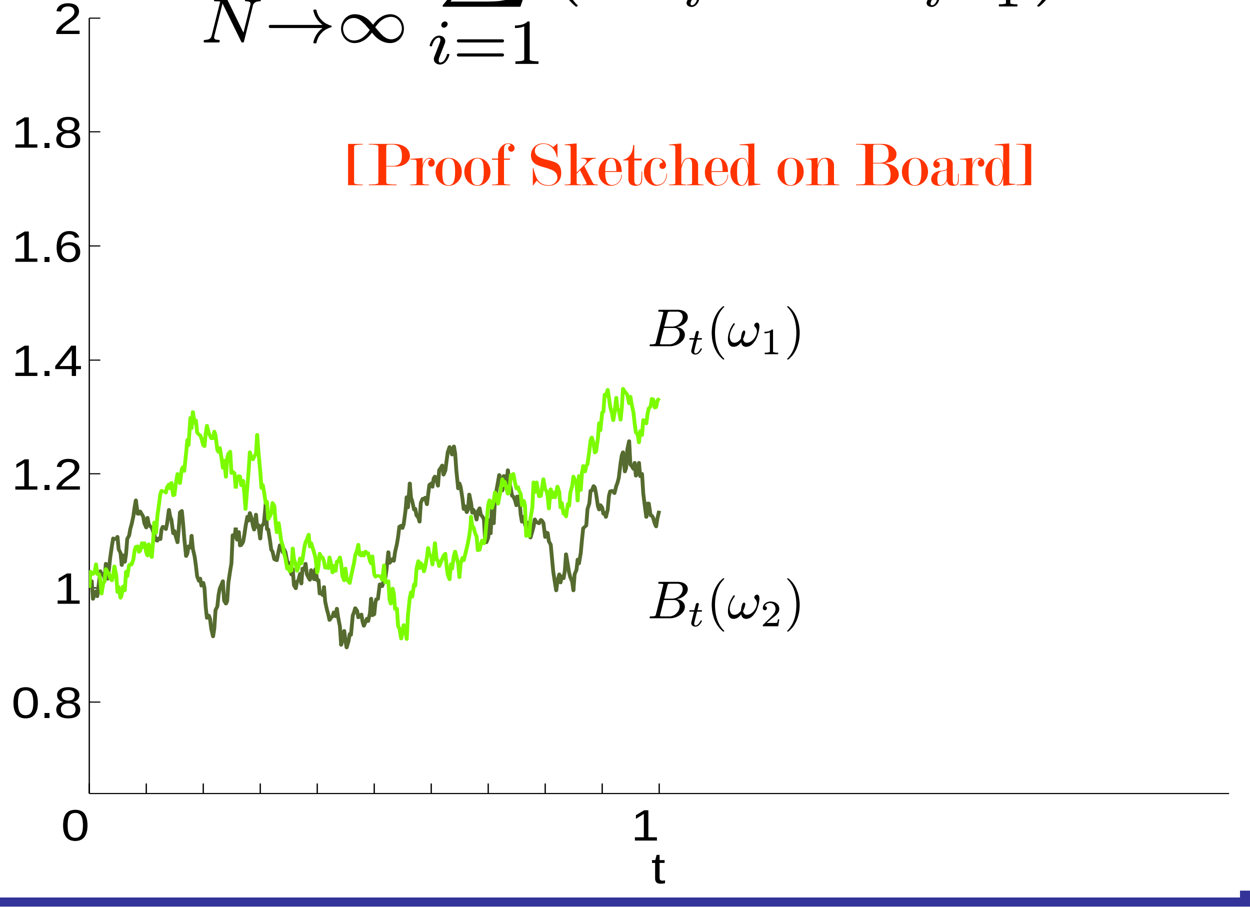
$$\sum_{i=1}^{k(t)} (B_{t_i} - B_{t_{i-1}})^2$$

$$k(t) = \max\{i : t_{i+1} \leq t\}$$

Different Paths, Same Limit

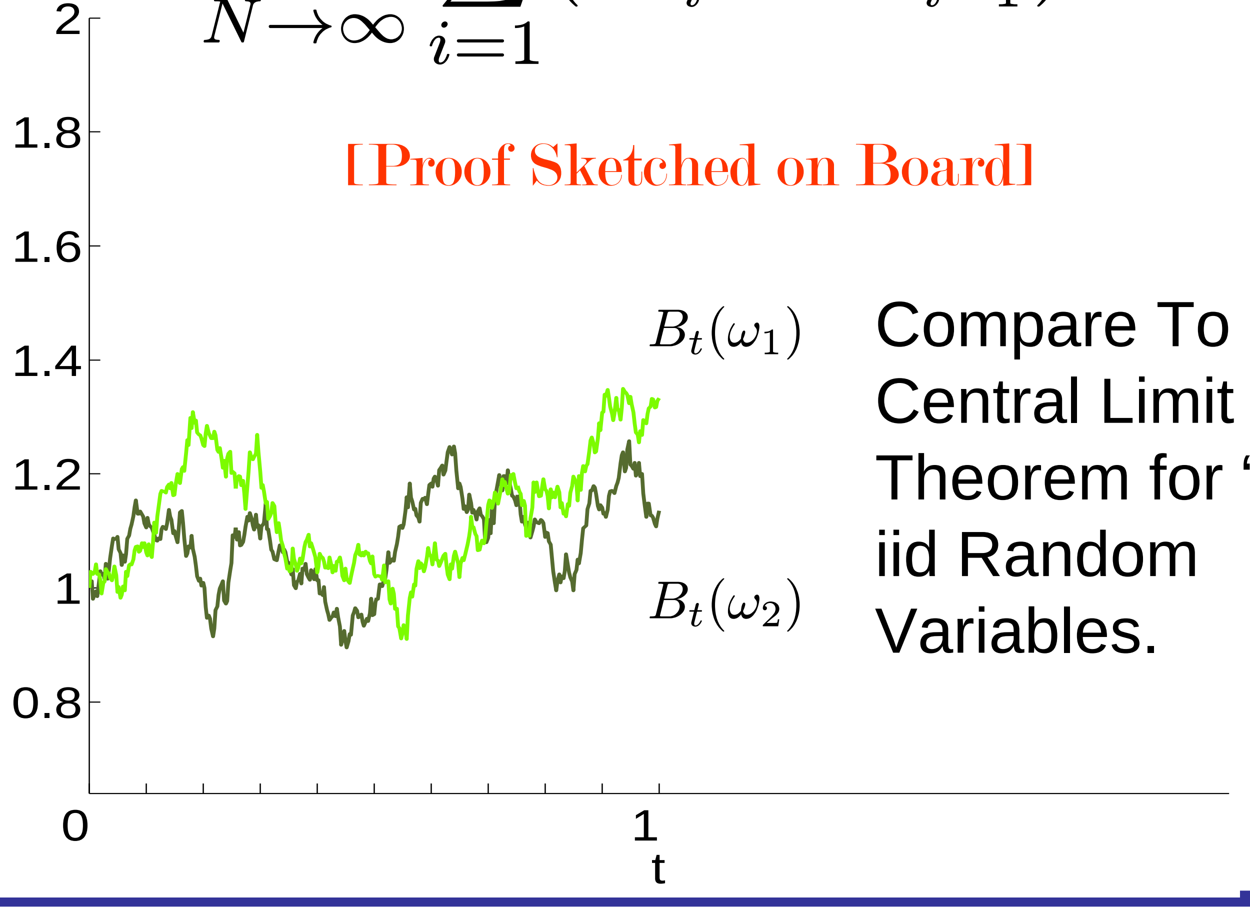
$$\lim_{N \to \infty} \sum_{i=1}^{k(t)} (B_{t_i} - B_{t_{i-1}})^2 = t$$

[Proof Sketched on Board]



Different Paths, Same Limit

$$\lim_{N\to\infty}\sum_{i=1}^{k(t)}(B_{t_i}-B_{t_{i-1}})^2=t$$

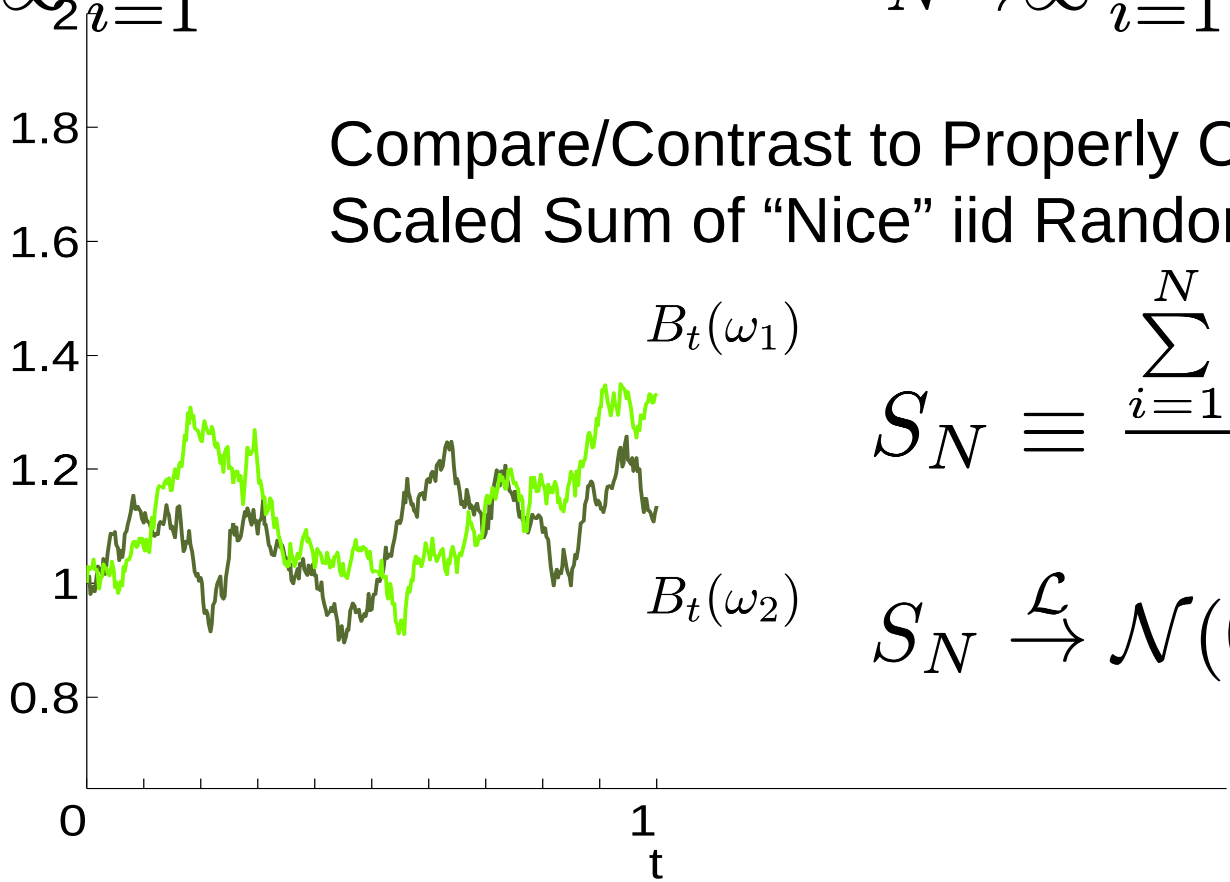


[Proof Sketched on Board]

Compare To Central Limit Theorem for “Nice” iid Random Variables.

Different Paths, Same Limit

$$\lim_{N\to\infty}\sum_{i=1}^{N}(B_{t_i}-B_{t_{i-1}})^2 \equiv \lim_{N\to\infty}\sum_{i=1}^{N} Y_i = T$$



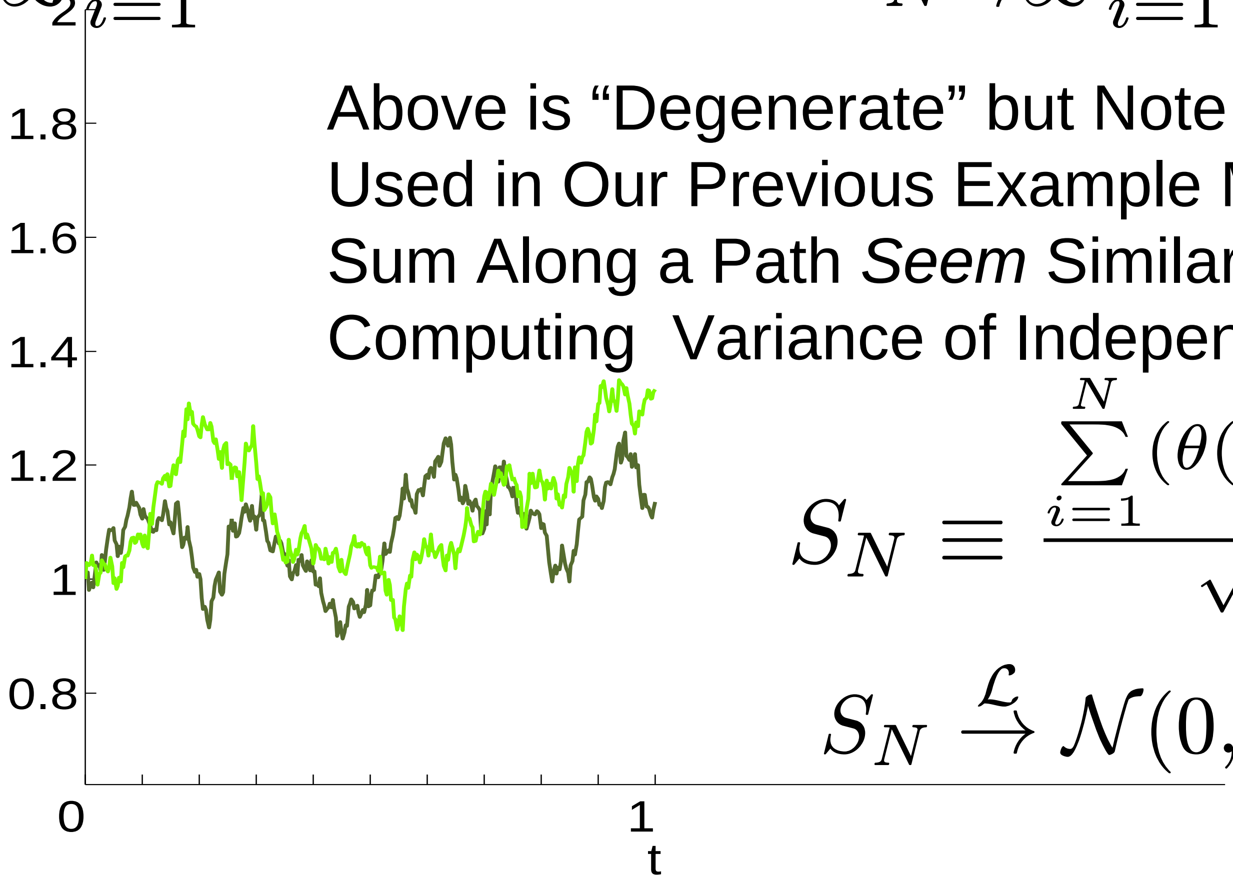
Compare/Contrast to Properly Centered and Scaled Sum of “Nice” iid Random Variables:

$$S_N \equiv \frac{\sum_{i=1}^{N}(\theta(\omega_i)-\mu)}{\sqrt{N}}$$

$$S_N \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

Continuity in Time

$$\lim_{N\to\infty}\sum_{i=1}^{N}(B_{t_i}-B_{t_{i-1}})^2 \equiv \lim_{N\to\infty}\sum_{i=1}^{N} Y_i = T$$



Above is "Degenerate" but Note Time Scaling Used in Our Previous Example Makes Infinite Sum Along a Path *Seem* Similar to Computing Variance of Independent RV's

$$S_N \equiv \frac{\sum_{i=1}^{N}(\theta(\omega_i)-\mu)}{\sqrt{N}}$$

$$S_N \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$$

B.M. Paths Do Not Have Finite Variation

$$\text{Let } 0 < t \leq T$$

$$\pi_n = \{0 \leq t_1 \leq \ldots t_i \leq \ldots \leq t_n = t\}$$

$$\lim_{n\to\infty} \max_{i<n}(t_{i+1} - t_i) = 0$$

$$\mathcal{P} \equiv \text{All Finite Partitionsof } [0, t]$$

B.M. Paths Do Not Have Finite Variation

$$\mathcal{V}_{[0,t]}(\omega) = \sup_{\pi \in \mathcal{P}} \sum_{t_i \in \pi} |B_{t_{i+1}} - B_{t_i}|$$

Suppose That: $\mathcal{P}(\mathcal{V}_{[0,t]} < \infty) > 0$

Then:

$$t = \lim_{N \to \infty} \sum_{i=1}^{N} (B_{t_i} - B_{t_{i-1}})^2$$
$$\leq \lim_{N \to \infty} \sup_{t_i \in \pi_N} |B_{t_i} - B_{t_{i-1}}| \sum_{i=1}^{N} |B_{t_i} - B_{t_{i-1}}|$$

B.M. Paths Do Not Have Finite Variation

$$t = \lim_{N\to\infty} \sum_{i=1}^{N} (B_{t_i} - B_{t_{i-1}})^2$$

$$\leq \lim_{N\to\infty} \sup_{t_i \in \pi_N} |B_{t_i} - B_{t_{i-1}}| \sum_{i=1}^{N} |B_{t_i} - B_{t_{i-1}}|$$

$$\leq \lim_{N\to\infty} \sup_{t_i \in \pi_N} |B_{t_i} - B_{t_{i-1}}| \mathcal{V}_{[0,t]}$$

$$= 0$$

Tends to Zero Due to a.s. uniform continuity of B.M. Paths

t > 0, so Probability of Finite Variation Must be Zero

Part II

Stochastic Integration and Quadratic Variation of SDEs

Stochastic Integration

$$\mathcal{P}(\mathcal{V}_{[0,t]} < \infty) = 0$$

$$X(\omega, t) = \int_0^t b(\omega, s)ds + \int_0^t \sigma(\omega, s)dB_s$$

"Finite Variation Term" "Infinite Variation Term"

Stochastic Integration

$$\mathcal{P}(\mathcal{V}_{[0,t]} < \infty) = 0$$

$$X(\omega,t) = \int_0^t b(\omega,s)ds + \int_0^t \sigma(\omega,s)dB_s$$

"Infinite Variation Term" Complicates
Interpretting Limit as Lebesgue Integral

$$\sum \sigma\big(X_{t_i}(B_{t_i})\big)|B_{t_{i+1}} - B_{t_i}|$$

Stochastic Differential Equations (SDEs)

$$X(\omega, t) = \int_0^t b(\omega, s)ds + \int_0^t \sigma(\omega, s)dB_s$$

Written in Abbreviated Form:

$$dX(\omega, t) = b(\omega, t)dt + \sigma(\omega, t)dB_t$$

Stochastic Differential Equations Driven by B.M.

$$dX(\omega,t) = b(\omega,t)dt + \sigma(\omega,t)dB_t$$

Adapted and Measurable Processes

$$dX_t = b_t dt + \sigma_t dB_t$$

I will use this shorthand for the above (which is itself shorthand for stochastic integrals…)

Stochastic Differential Equations Driven by B.M.

$$dX_t = b_t dt + \sigma_t dB_t$$

General Ito Process

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dB_t$$

Diffusion Process

$$dX_t = 0dt + 1dB_t$$

Repackaging Brownian Motion

Quadratic Variation for General SDEs (Jump Processes Readily Handled)

$$RV_{\pi_N}(t) \equiv \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})^2$$

"Realized Variation"

Quadratic Variation: Defined if the limit taken over any partition exists with the mesh going to zero.

$$:= \{ \mathrm{V}_t \}$$

Stochastic Integration

$$Y(\omega, t) = \int_0^t b'(\omega, s)ds + \int_0^t \sigma'(\omega, s)dX_s$$

"Finite Variation Term"

"Infinite Variation Term"

Quadratic Variation of "X" Can Be Used to Define Other SDEs (A "Good Stochastic Integrator")

Quadratic Variation Comes Entirely From Stochastic Integral

$$X(\omega,t)=\int_0^t b(\omega,s)ds+\int_0^t \sigma(\omega,s)dB_s$$

i.e. the drift (or term in front of “ds” makes) no contribution to the quadratic variation

[quick sketch of proof on board]

Stochastic Integration

$$\mathcal{P}(\mathcal{V}_{[0,t]} < \infty) = 0$$

$$X(\omega, t) = \int_0^t b(\omega, s) ds + \int_0^t \sigma(\omega, s) dB_s$$

"Infinite Variation Term" Can Now Be Dealt With If One Uses Q.V.

$$\sum \sigma\big(X_{t_i}(B_{t_i})\big) |B_{t_{i+1}} - B_{t_i}|$$

Quadratic Variation and “Volatility”

$$X_t = \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

Typical Notation for QV

$$[X, X]_t \equiv V_t = \int_0^t \sigma_s^2 dt$$

Quadratic Variation and "Volatility"

For Diffusions

$$[X,X]_t \equiv V_t = \int_0^t \sigma_s^2 dt$$

More Generally for Semimartingales
(Includes Jump Processes)

$$[X,X]_t := X_t^2 - 2\int_0^t X_{s-} dX_s$$

Part III

Demonstration of How Results Help Understand ”Realized Variation” Discretization Error and Idea Behind “Two Scale Realized Volatility Estimator”

Most People and Institutions Don't Sample Functions (Discrete Observations)

$$RV_{\pi_N}(t) \equiv \sum_{i=1}^{N} (X_{t_i} - X_{t_{i-1}})^2$$

Assuming Process is a Genuine Diffusion, How Far is Finite N Approximation From Limit?

$$[X, X]_t$$

Discretization Errors and Their Limit Distributions

$$N^{1/2}\Big(RV_{\pi_N}(t) - [X,X]_t\Big)|\sigma_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, C\int_0^t \sigma_s^4 ds)$$

Paper Below Derived Explicit Expressions for Variance for Fairly General SDEs Driven by B.M.

Barndorff-Nielsen, O. & Shepard, N. (2003) *Bernoulli* **9** 243-265.

Discretization Errors and Their Limit Distributions

$$N^{1/2}\left(RV_{\pi_N}(t) - [X,X]_t\right)|\sigma_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, C\int_0^t \sigma_s^4 ds)$$

Their paper goes over some nice classic moment generating function results for wide class of spot volatility processes (and proves things beyond QV)

Barndorff-Nielsen, O. & Shepard, N. (2003) *Bernoulli* **9** 243-265.

Discretization Errors and Their Limit Distributions

$$\delta \equiv \frac{t}{N}$$

One Condition That They Assume:

$$\lim_{N \to 0} \delta^{1/2} \sum_{i=1}^{N} |\sigma^2_{t_i+\delta} - \sigma^2_{t_i}| = 0$$

Barndorff-Nielsen, O. & Shepard, N. (2003) *Bernoulli* **9** 243-265.

Discretization Errors and Their Limit Distributions

$$\delta \equiv \frac{t}{N}$$

Good for Many Stochastic Volatility Models … BUT Some Exceptions are Easy to Find (See Board)

$$\lim_{N \to 0} \delta^{1/2} \sum_{i=1}^{N} |\sigma^2_{t_i+\delta} - \sigma^2_{t_i}| = 0$$

Barndorff-Nielsen, O. & Shepard, N. (2003) *Bernoulli* **9** 243-265.

What Happens When One Has to Deal with "Imperfect" Data?

Suppose "Price" Process Is Not Directly Observed But Instead One Has:

$$y_{t_i} = X_{t_1} + \epsilon_i$$

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

What Happens When One Has to Deal with "Imperfect" Data?

Due to Fact that Real World Data Does Not Adhere to Strict Mathematical Definition of a Diffusion

$$y_{t_i} = X_{t_1} + \epsilon_i$$

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

What Happens When One Has to Deal with “Imperfect” Data?

$$y_{t_i} = X_{t_1} + \epsilon_i$$

If “observation noise” sequence is i.i.d. (and independent of X) and it is very easy to see problem of estimating quadratic variation (see demostration on board).

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

What Happens When One Has to Deal with "Imperfect" Data?

$$y_{t_i} = X_{t_1} + \epsilon_i$$

If "observation noise" sequence is i.i.d. (and independent of X) and it is very easy to see problem of estimating quadratic variation (see demostration on board). Surprisingly These Stringent Assumption Model Many Real World Data Sets (e.g. See Results from Lecture 1)

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

Simple Case "Bias" Estimate

Compute Realized Variation of Y and Divide by 2N

Then Obtain (Biased) Estimate by Subsampling

Finally Aggregate Subsample Estimates and Remove Bias

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

Simple Case “Bias Corrected” Estimate

$$\widehat{[X,X]}_T = [Y,Y]_T^{(\mathrm{avg})} - \frac{N_s}{N}[Y,Y]_T^{(\mathrm{all})}$$

Result Based on “Subsampling”

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

Chasing A (Time Series) Dream

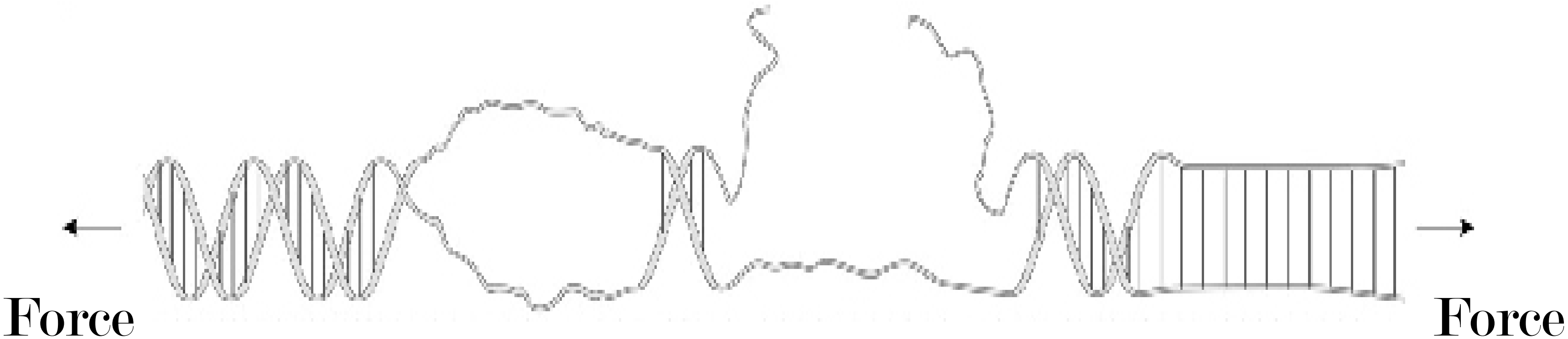
Obtaining a "Consistent" Estimate with Asymptotic Error Bounds

$$N^{1/6}(\widehat{[X,X]}_T - [X,X]_T)|\sigma_t \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, C_1(\mathrm{VAR}(\epsilon))^2 + C_2\int_0^T \sigma_s^4 ds\right)$$

Result Above Valid for Simple Observation Noise but Results (by Authors Below) Have Been Extended to Situations More General Than iid Case.

Zhang, L.. Mykland, P. & Ait-Sahalia,Y. (2005) *JASA* **100** 1394-1411.

DNA Melting



Tensions Important in Fundamental Life Processes Such As: DNA Repair and DNA Transcription

Single-Molecule Experiments Not Just a Neat Toy:

They Have Provided Insights Bulk Methods Cannot

Time Dependent Diffusions

Stochastic Differential Equation (SDE):

$$dz_t = b(z_t, t; \Theta)dt + \sigma(z_t; \Theta)dW_t$$

$$y_{t_i} = z_{t_i} + \epsilon_{t_i}$$ "Measurement" Noise

For Frequently Sampled Single-Molecule Experimental Data, "Physically Uninteresting Measurement Noise" Can Be Large Component of Signal.

Find/Approximate: $p(z_t, |y_s; \Theta)$

"Transition Density" / Conditional Probability Density

A Word from the Wise (Lucien Le Cam)

- Basic Principle 0: Don't trust any principle
- Principle 1: Have clear in your mind what it is you want to estimate

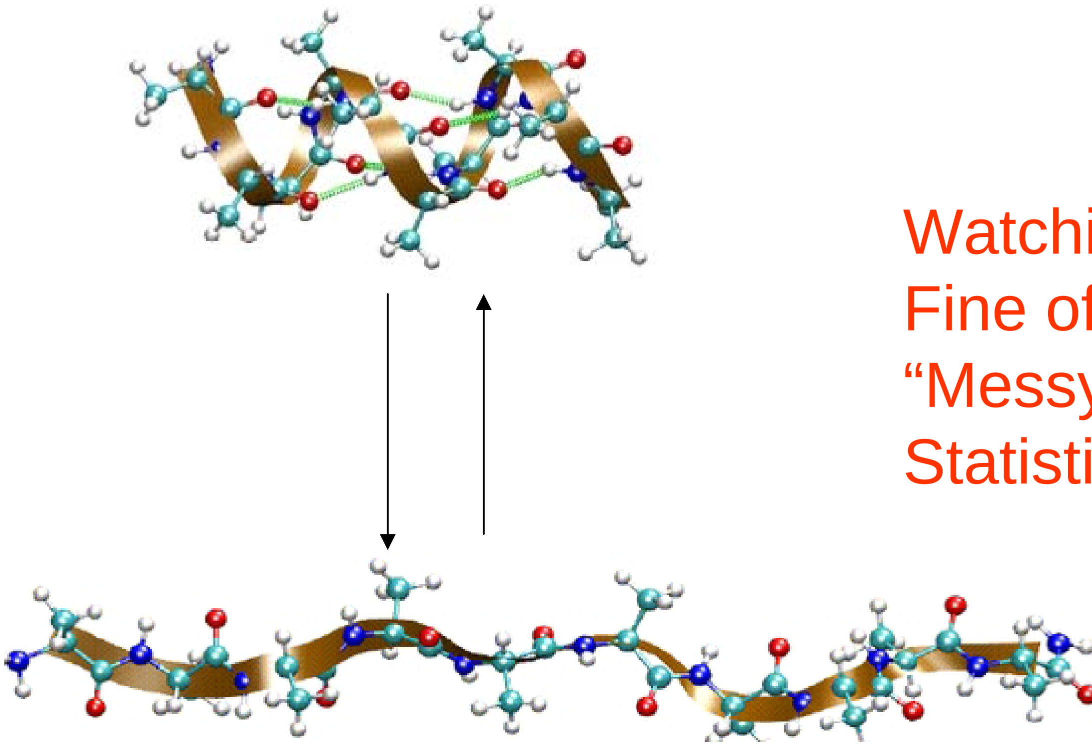
…

- Principle 4: If satisfied that everything is in order, try first a crude but reliable procedure to locate the general area your parameters lie.
- Principle 5: Having localized yourself by (4), refine the estimate using some of your theoretical assumptions, being careful all the while not to undo what you did in (4).

…

Le Cam L.. (1990) *International Statistics Review* **58** 153-171.

Example of Relevance of Going Beyond iid Setting

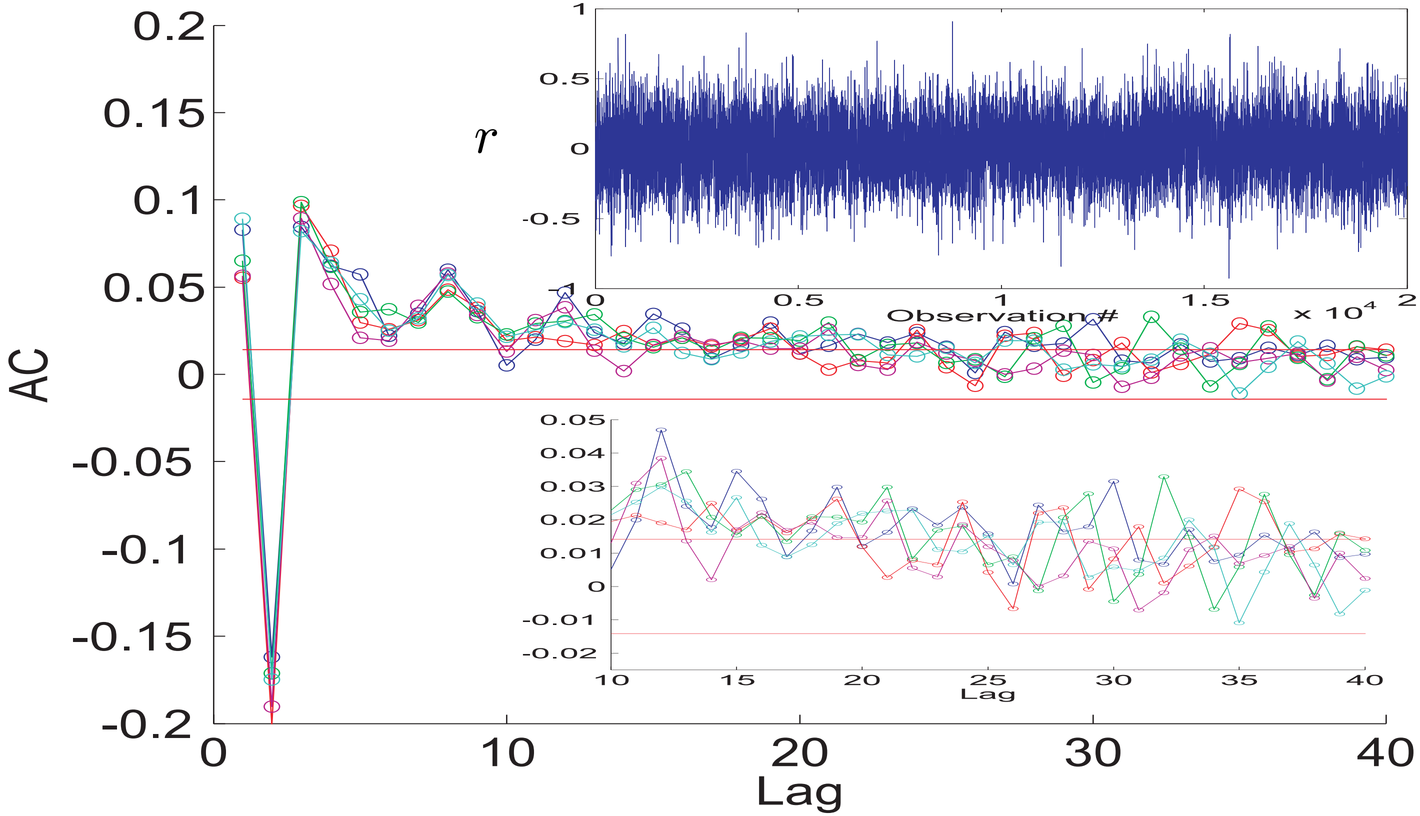


Watching Process at Too Fine of Resolution Allows "Messy" Details to Be Statistically Detectable

Calderon, *J. Chem. Phys.* (2007).

Example of Relevance of Going Beyond iid Setting

$$r \equiv X_n - \mathrm{E}^{\mathbb{P}_{\theta_i}}[X_n | X_{n-1}]$$



Calderon, *J. Chem. Phys.* (2007).

Lecture IV: A Selected Review of Some Recent Goodness-of-Fit Tests Applicable to Levy Processes

Christopher P. Calderon

Rice University / Numerica Corporation

Research Scientist

Outline

I Compare iid Goodness of Fit Problem to Time Series Case

II Review a Classic Result Often Attributed to Rosenblatt (The Probability Integral Transform)

III Highlight Hong and Li's "Omnibus" Test

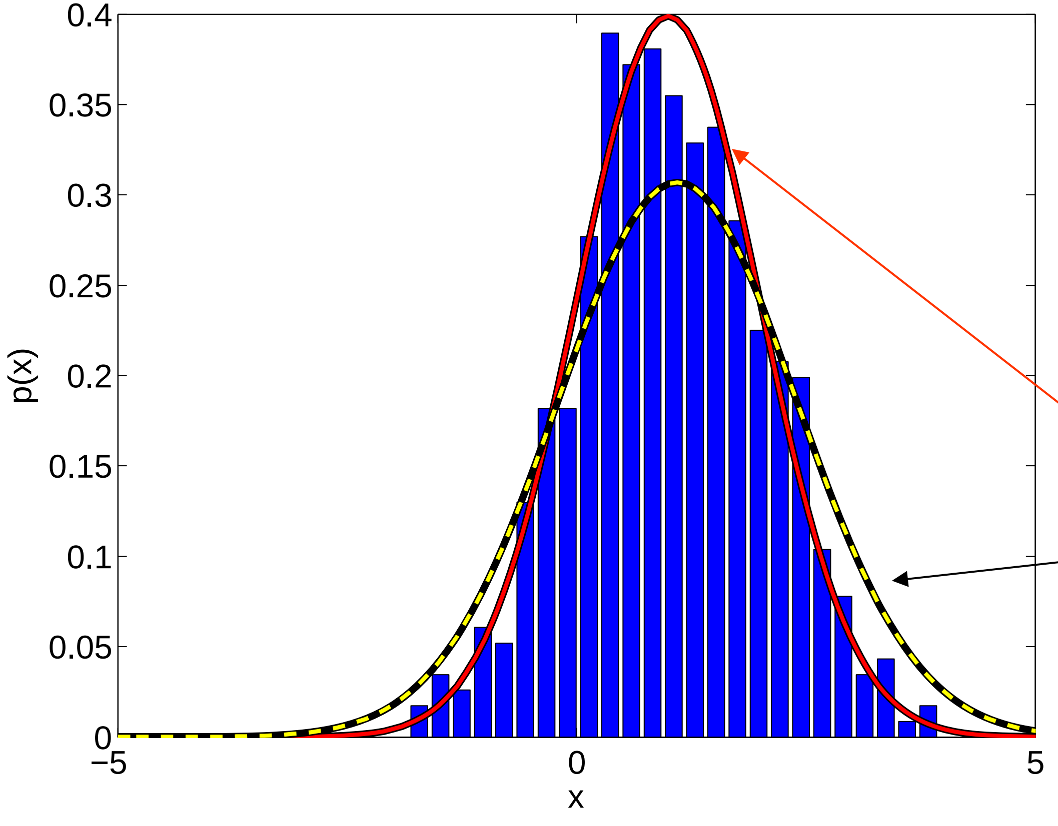
IV Demonstrate Utility and Discuss Some Other Recent Approaches and Open Problems

Outline

Primary References (Item II & III):

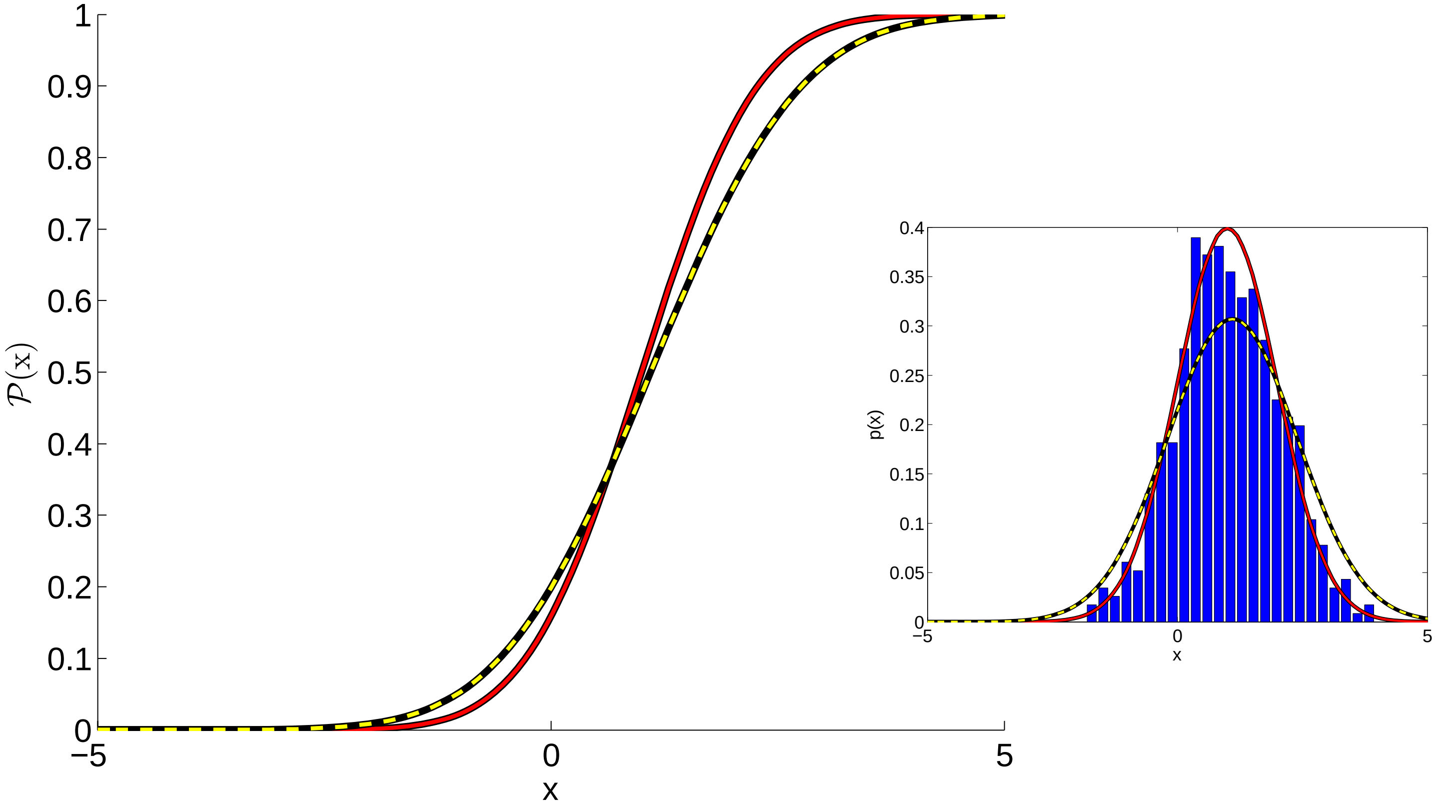
Diebold, F. Hahn J., & Tay, A. (1999) *Review of Economics and Statistics* **81** 661-663.
Hong, Y. and Li, H. (2005) *Review of Financial Studies* **18** 37-84.

Goodness-of-Fit Issues .

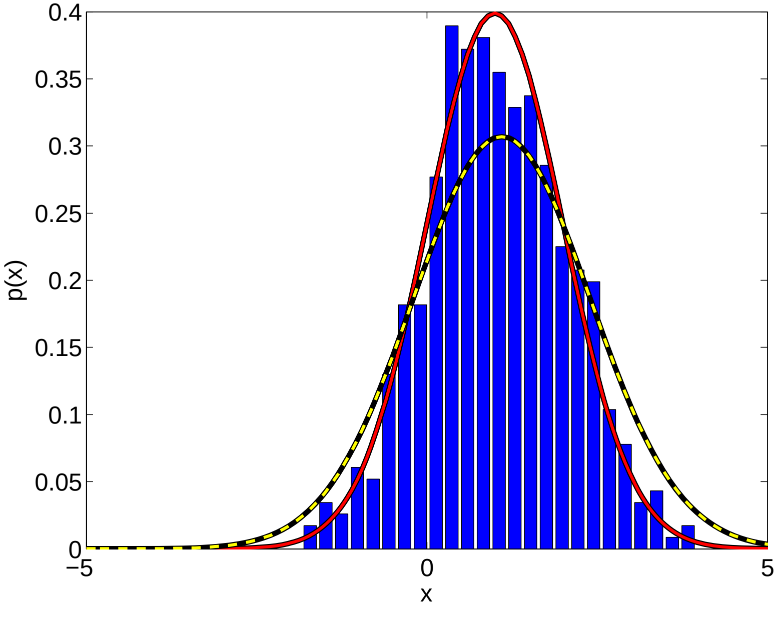


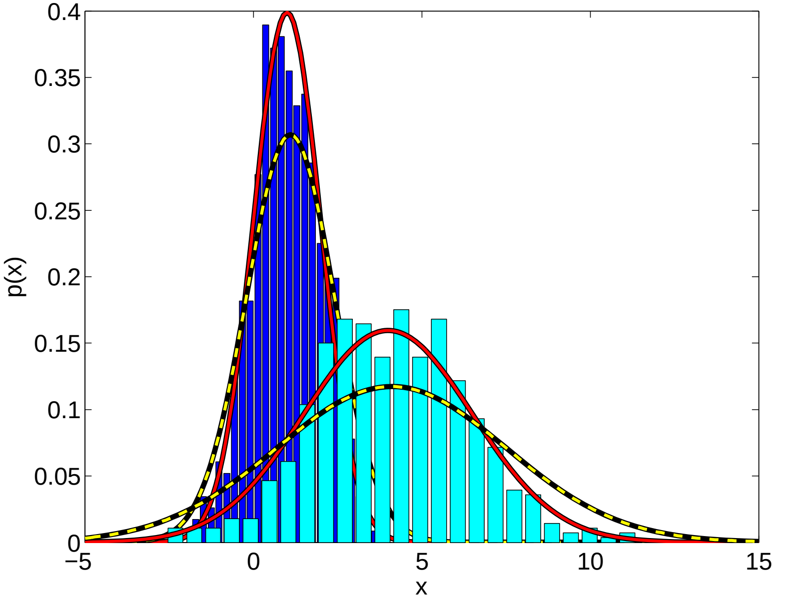
True and Parametric Density

Goodness-of-Fit in Random Variable Case (No Time Evolution)

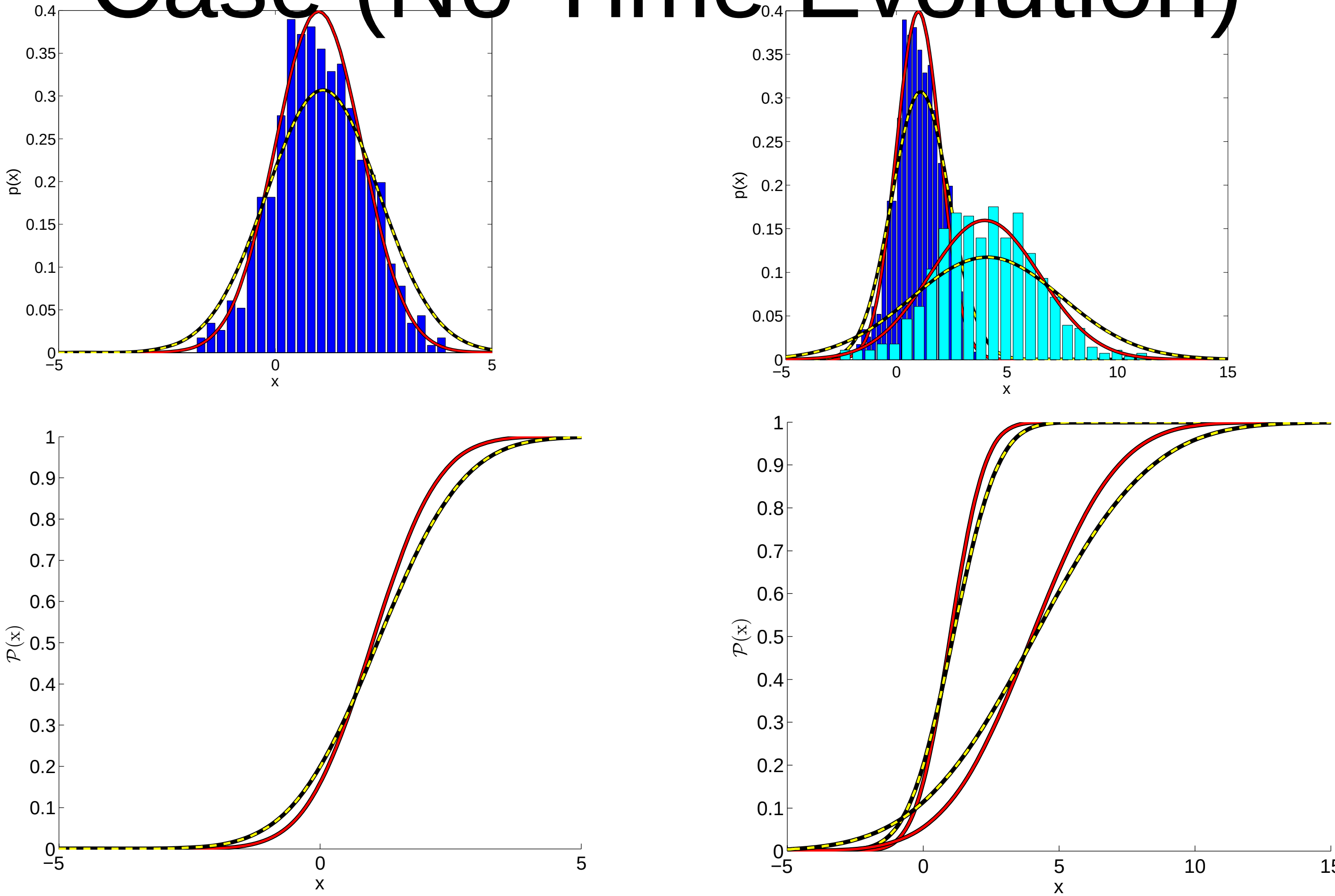


Goodness-of-Fit Issues

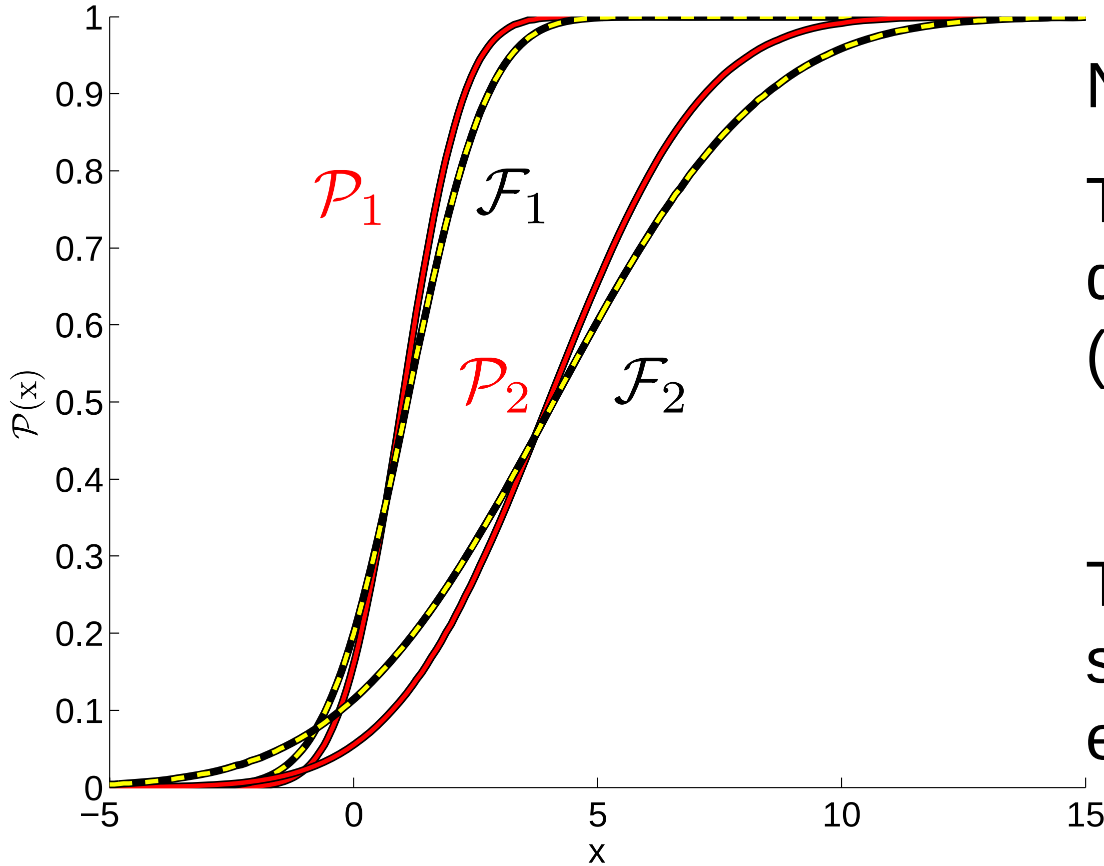




Goodness-of-Fit in Random Variable Case (No Time Evolution)



Goodness-of-Fit in Time Series



Note:

Truth and approximate distribution changing with (known) time index

Time series only get “one” sample from each evolving distribution

Time Series: Residual and Extensions

$$r_i \equiv X_i - \mathbb{E}^{\hat{\theta}}[X_i | X_{i-1}]$$

Problem with Using Low Order Correlation Shown In Lecture 2.

Can we Make Use of All Moments?

Time Series: Residual and Extensions

$$r_i \equiv X_i - \mathbb{E}^{\hat{\theta}}[X_i|X_{i-1}]$$

Problem with Using Low Order
Correlation Shown In Lecture 2.

Can we Make Use of All Moments?

YES. With the Probability Integral
Transform (PIT)

$$Z_i = \int_{-\infty}^{X_i} p(X|X_{i-1};\theta_0)dX$$

Also Called Generalized Residuals

Test the Model Against Observed Data:

$$\{X_1, X_2, \ldots X_N\} \dashrightarrow \hat{\theta}$$

Time Series (Noisy, Correlated, and Possibly Nonstationary)

$$Z_i = G(X_i; X_{i-1}, \hat{\theta})$$

Probability Integral Transform / Rosenblatt Transform /"Generalized Residual"

If Assumed Model Correct, Residuals i.i.d. with Known Distribution::

Formulate Test Statistic::

$$Z_i \sim U[0,1] \dashrightarrow Q = H(\{Z_1, Z_2, \ldots Z_T\})$$

Simple Illustration of PIT

[Specialized to Markov Dynamics]

Start time series random variables characterized by : $X_n \sim \mathcal{P}(X_n|X_{n-1})$

Introduce (strictly increasing) transformation: $Z_n = h(X_n;\theta)$

Transformation introduces "new" R.V. => New distribution $Z_n \sim \mathcal{F}(Z_n|Z_{n-1};\theta)$

"Truth" density in "X" (NOTE: no parameter) $p(X_n|X_{n-1}) \equiv \frac{d\mathcal{P}(X_n|X_{n-1})}{dX_n}$

PIT transformation: $Z_n \equiv h(X_n) := \int\limits_{-\infty}^{X_n} f(X'|X_{n-1};\theta)dX'$

Want expression for "new" density in terms of "Z"

Monotone Function of X

$$\frac{d\mathcal{F}(Z_n|Z_{n-1};\theta)}{dZ_n} = \frac{d\mathcal{P}(X_n|X_{n-1})}{dX_n}\frac{dX_n}{dZ_n}$$

$$= \frac{d\mathcal{P}(X_n|X_{n-1})}{dX_n}\frac{1}{\frac{dZ_n}{dX_n}} = p(X_n|X_{n-1})\frac{1}{f(X_n|X_{n-1};\theta)} \stackrel{?}{=} 1$$

$$\hat{Q}(j) \equiv \Big(h(N-j)\hat{M}_1(j) - hA^0 \Big) / V_0^{\frac{1}{2}}$$

Location (Depends on Bandwidth)

Scale

$$\hat{M}_1(j) \equiv \int_0^1 \int_0^1 \Big(\hat{g}_j(z_1, z_2) - 1 \Big)^2 dz_1 dz_2.$$

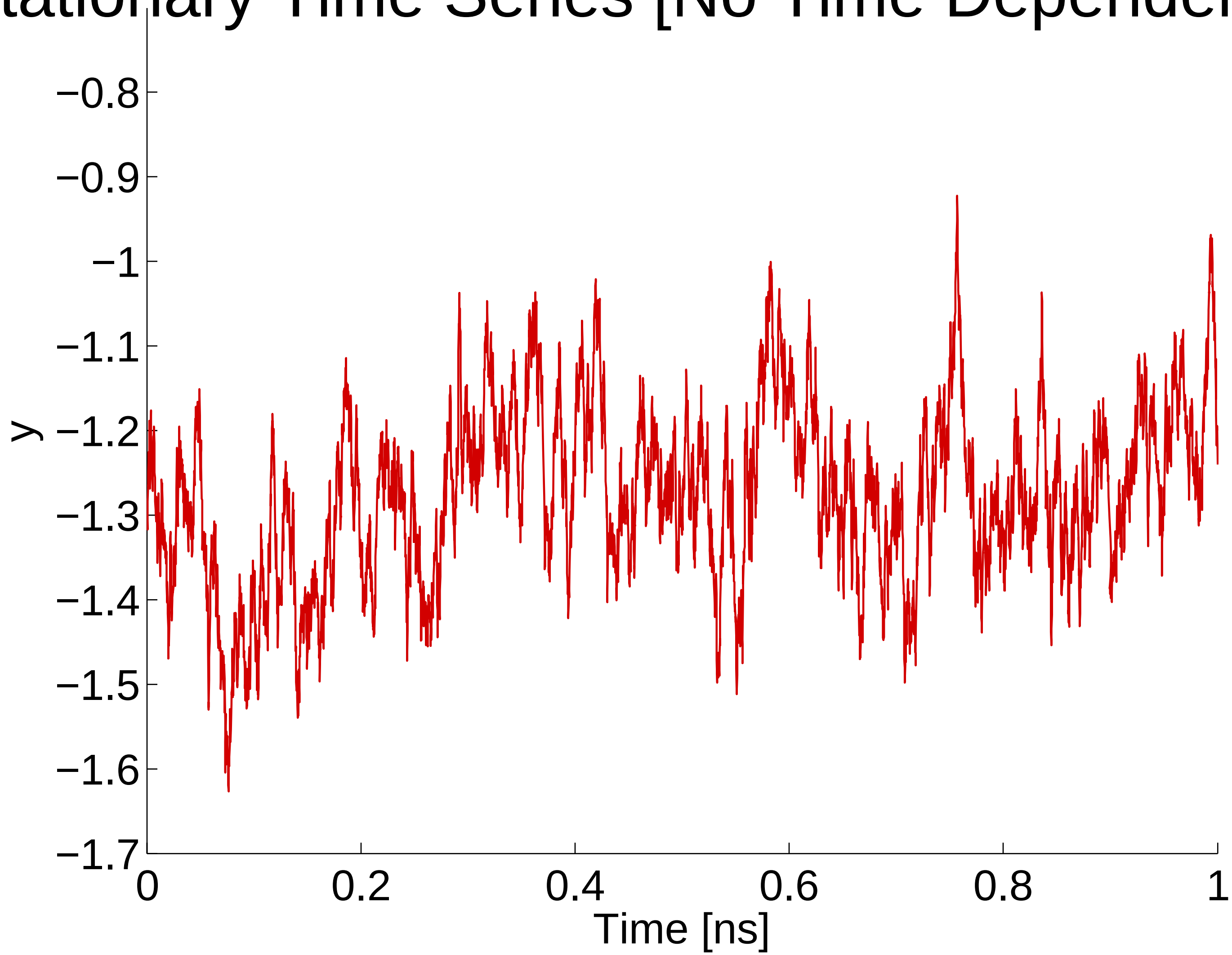
$$\hat{g}_j(z_1, z_2) \equiv \frac{1}{N-j} \sum_{\tau=j+1}^{N} K_h(z_1, \hat{Z}_\tau) K_h(z_2, \hat{Z}_{\tau-j}).$$

$$\hat{Q}(j) \equiv \Big(h(N-j)\hat{M}_1(j) - hA^0 \Big) / V_0^{\frac{1}{2}}$$

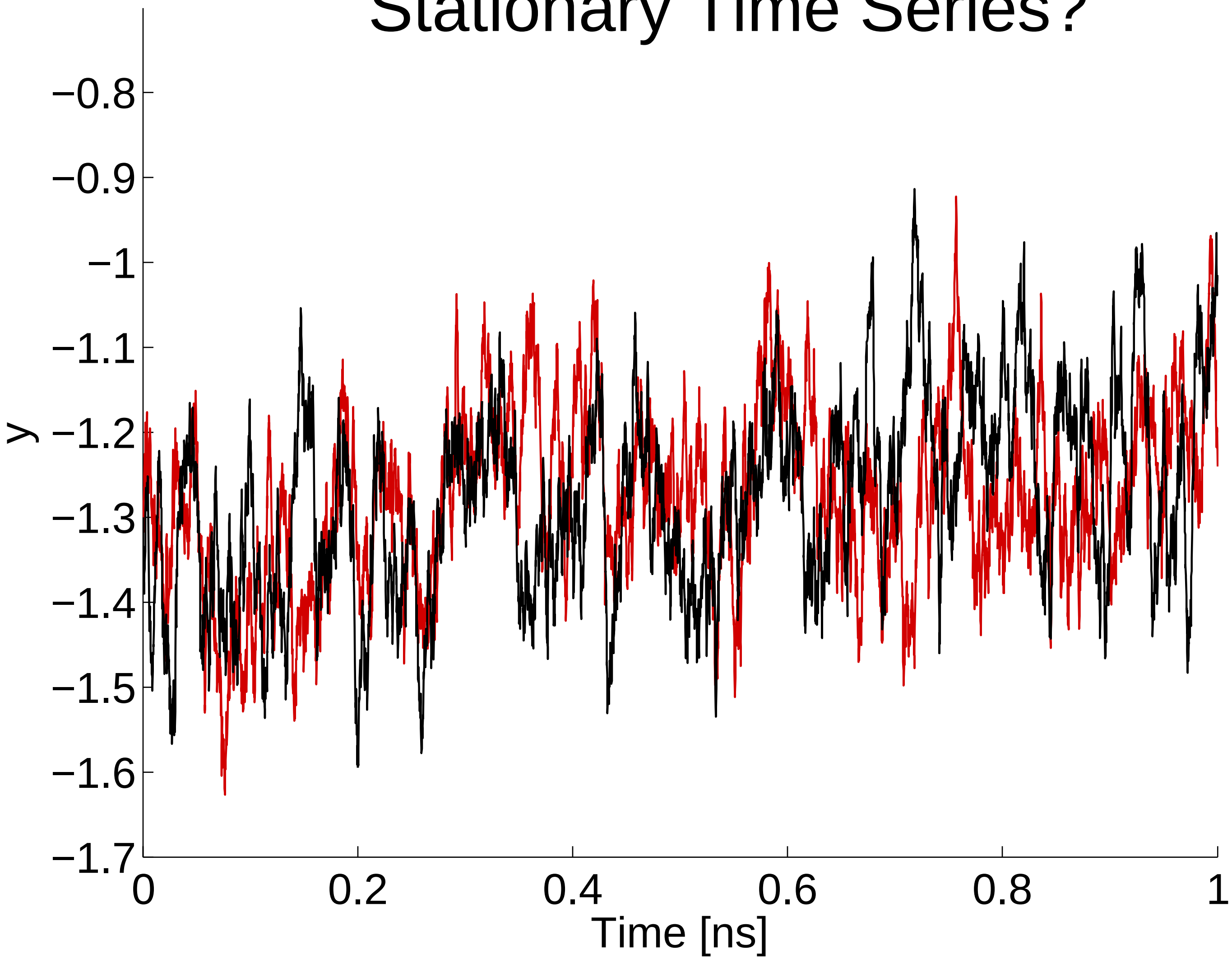
$$\hat{M}_1(j) \equiv \int_0^1 \int_0^1 \Big(\hat{g}_j(z_1, z_2) - 1 \Big)^2 dz_1 dz_2.$$

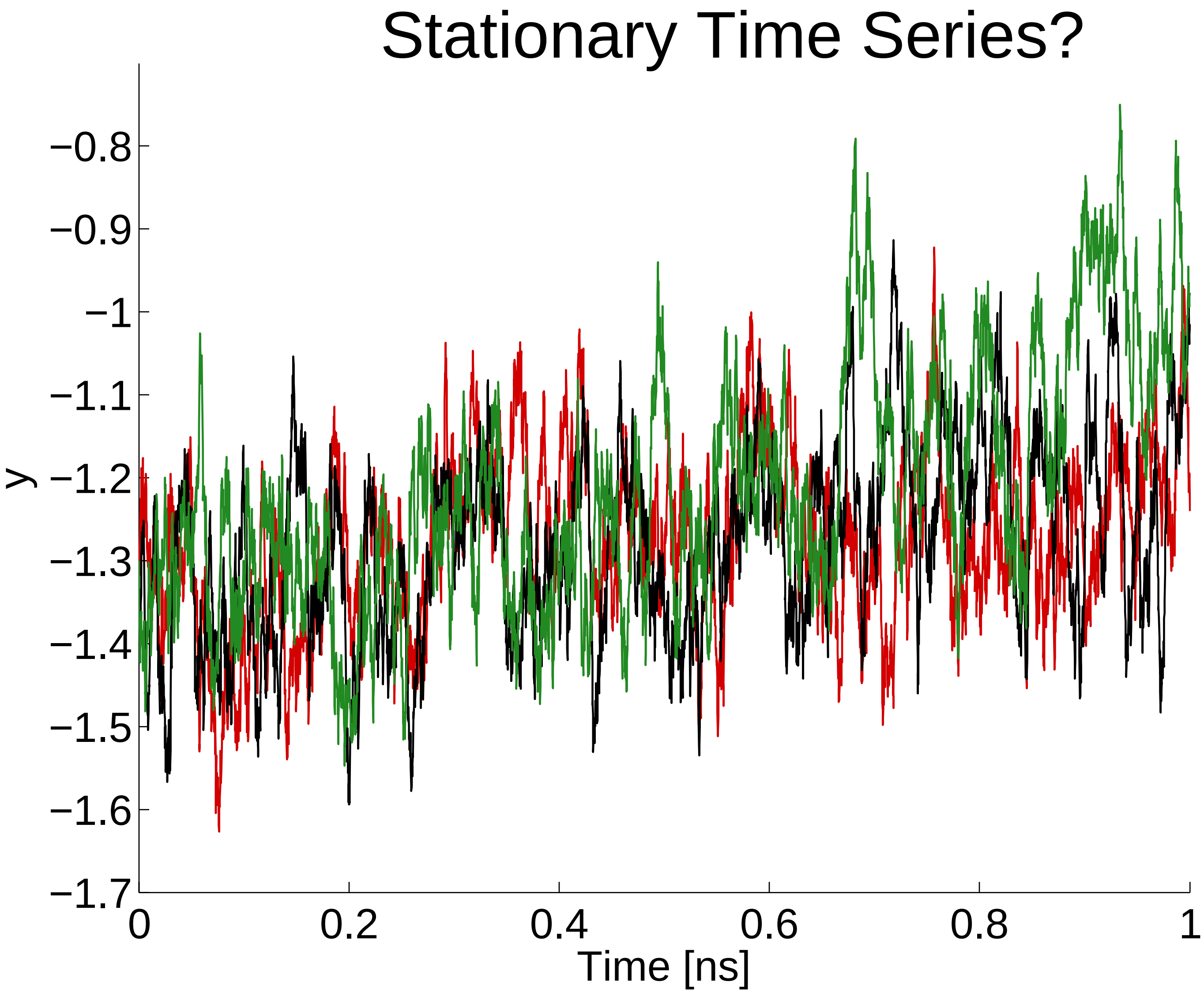
$$\hat{g}_j(z_1, z_2) \equiv \frac{1}{N-j} \sum_{\tau=j+1}^{N} K_h(z_1, \hat{Z}_\tau) K_h(z_2, \hat{Z}_{\tau-j}).$$

Stationary Time Series [No Time Dependent Forcing]



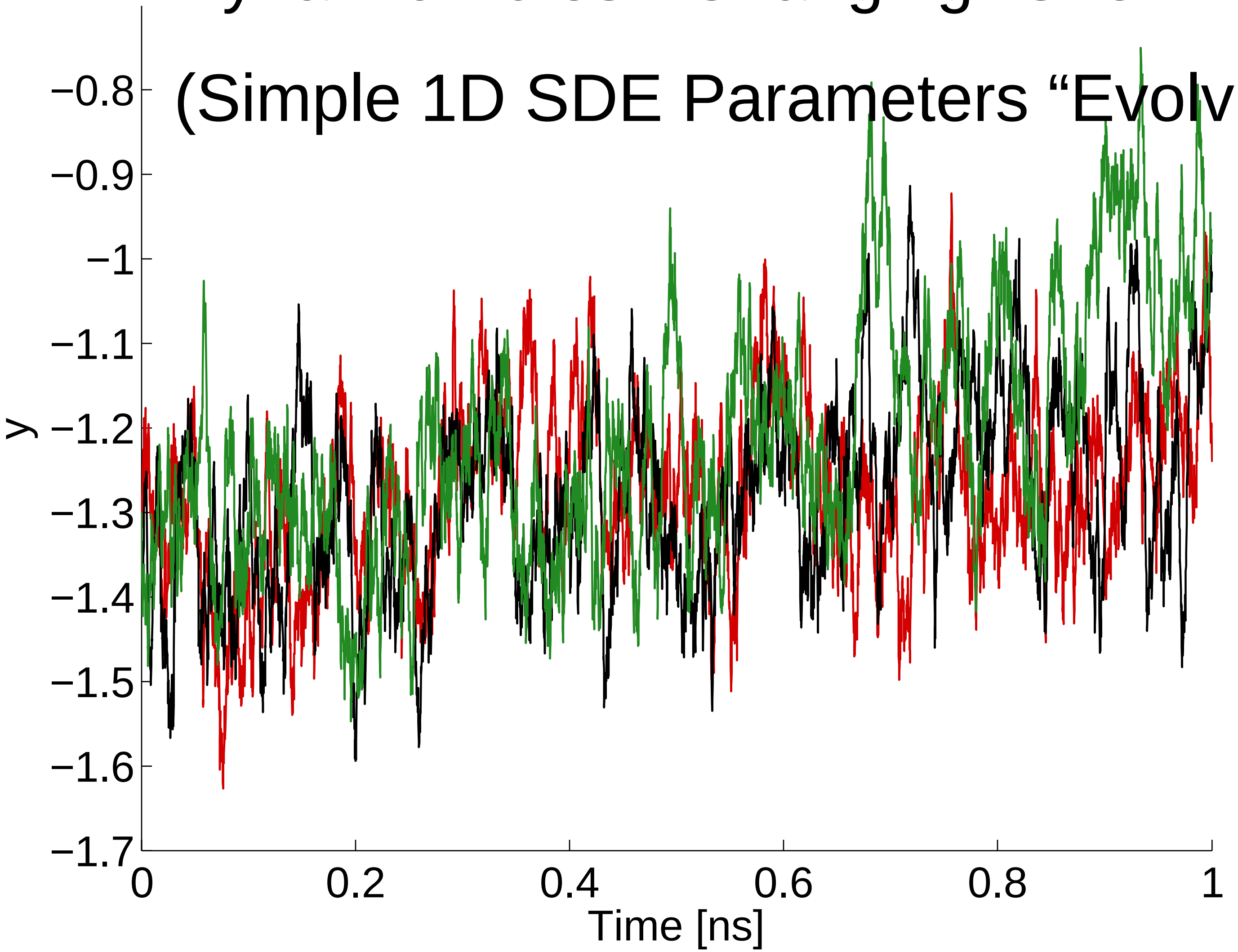
Stationary Time Series?




Stationary Time Series?
y
−0.8
−0.9
−1
−1.1
−1.2
−1.3
−1.4
−1.5
−1.6
−1.7
0
0.2
0.4
0.6
0.8
1
Time [ns]

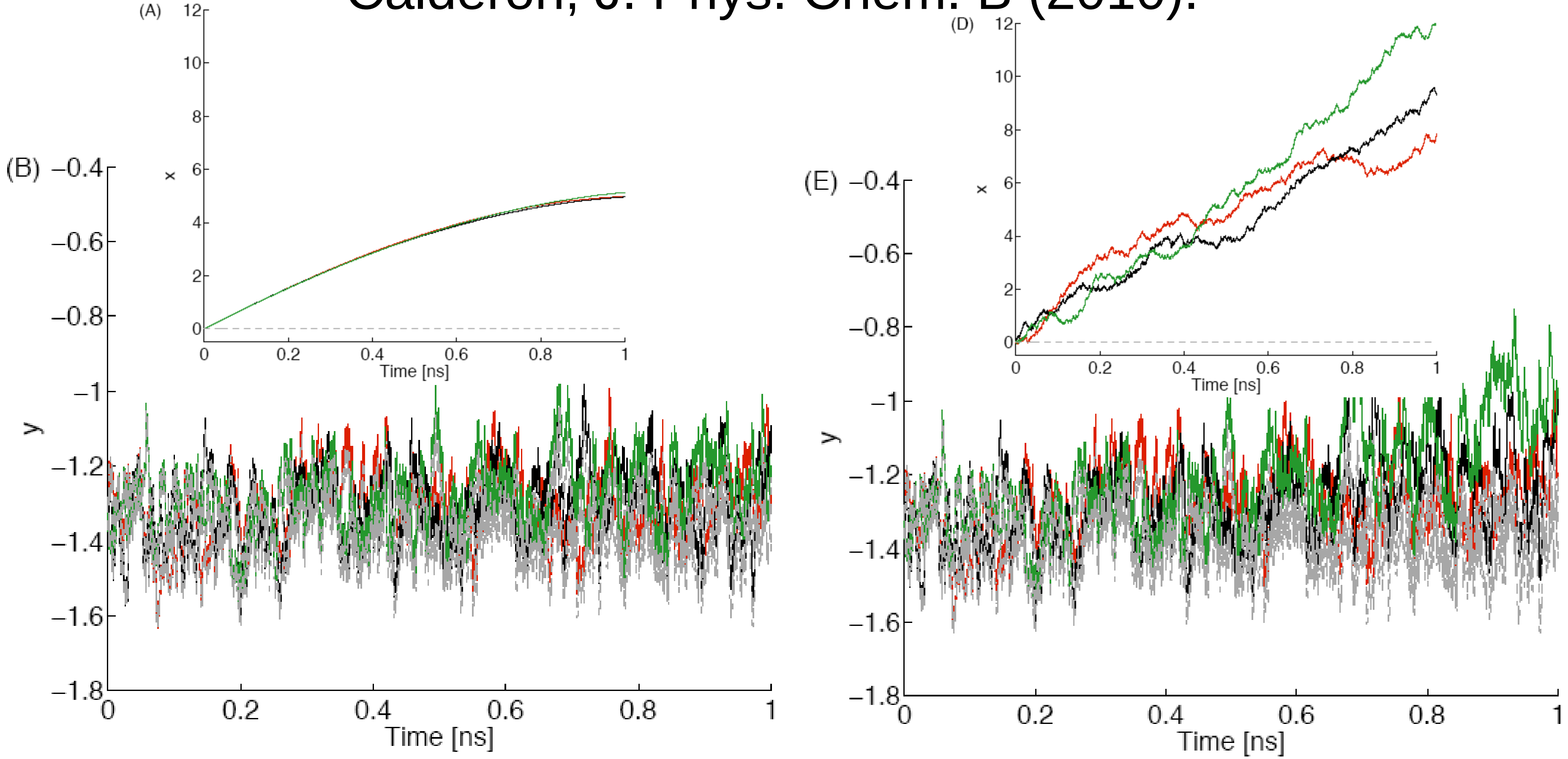
Dynamic Rules “Changing” Over Time

(Simple 1D SDE Parameters “Evolving”)

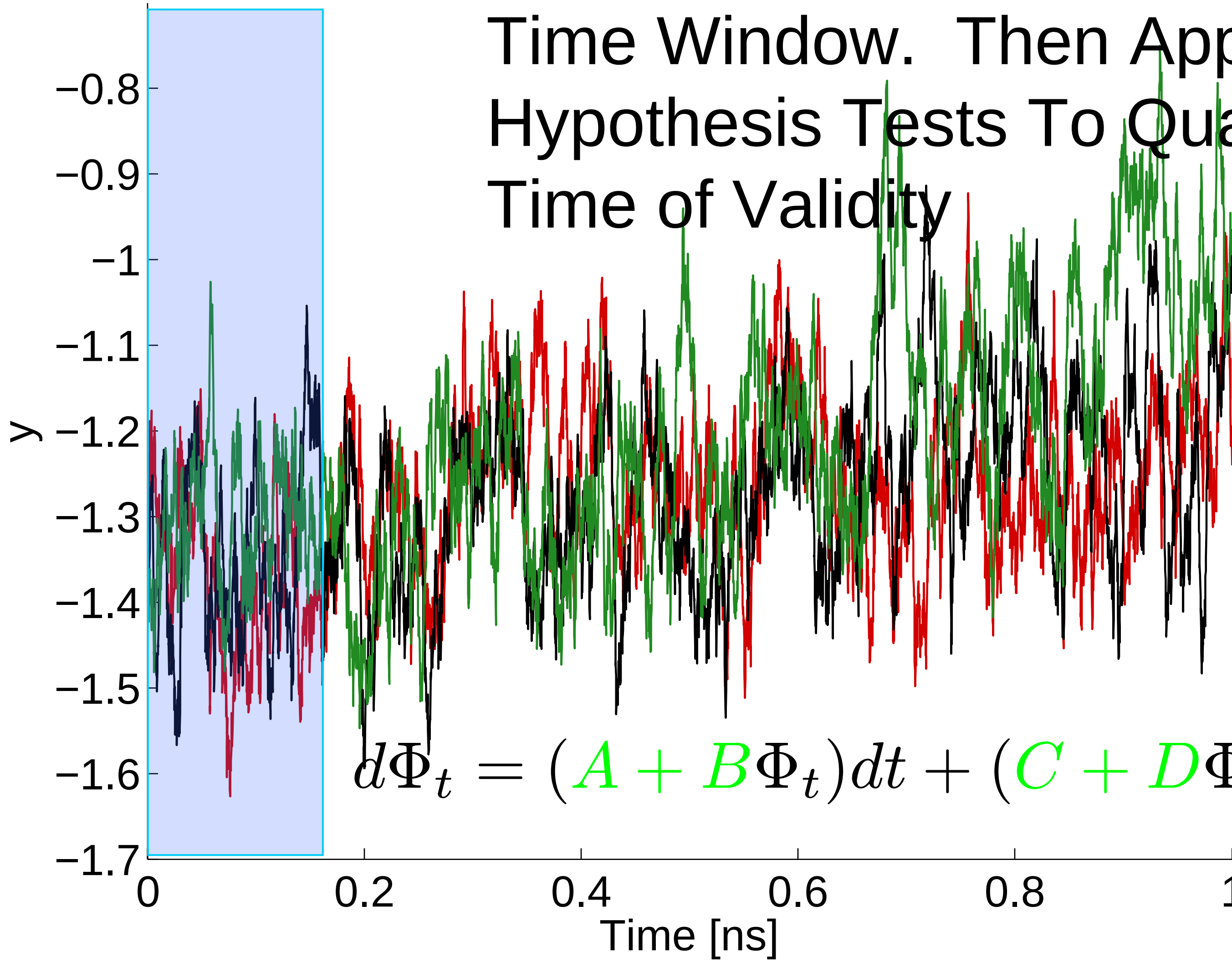


“Subtle” Coupling

Calderon, J. Phys. Chem. B (2010).

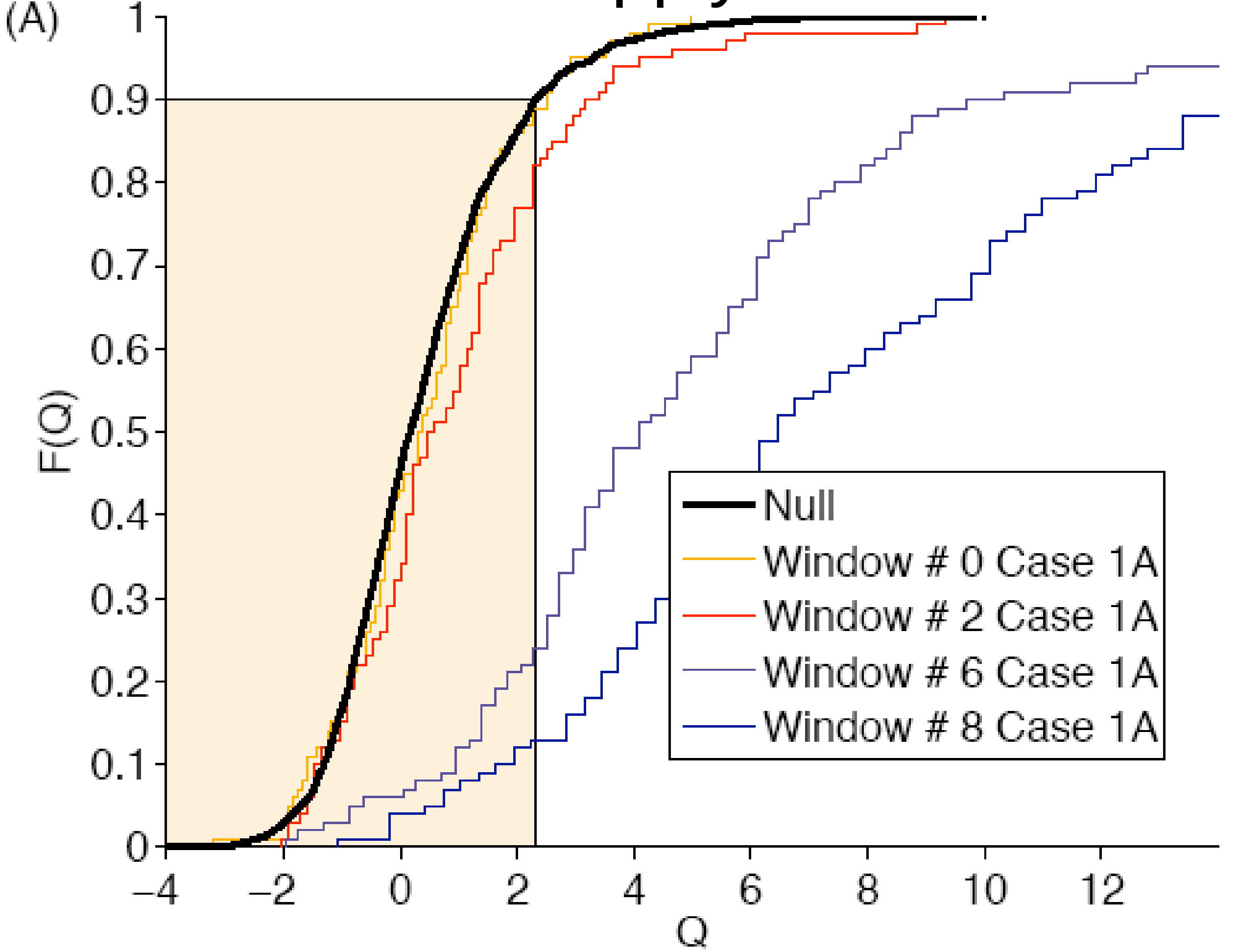


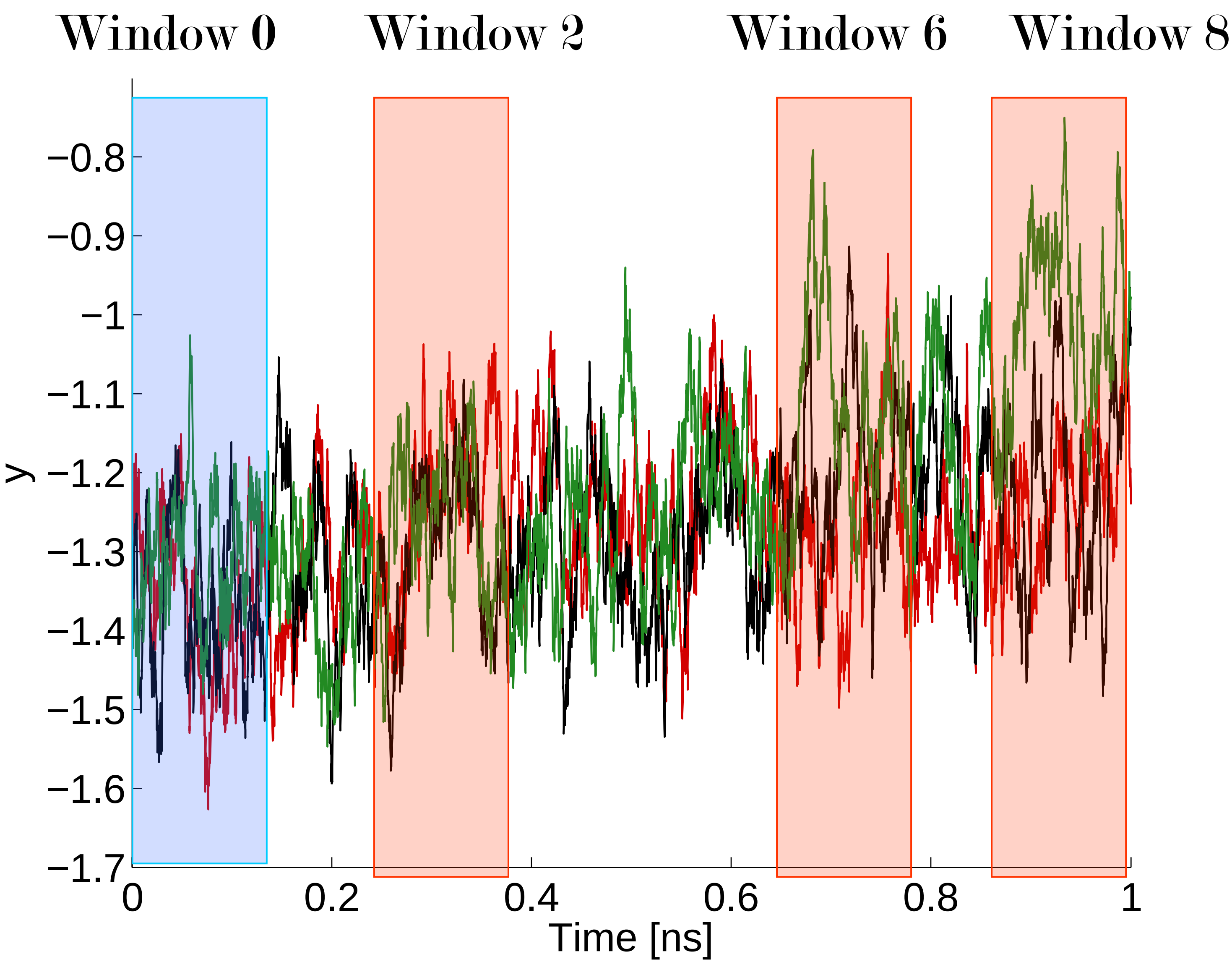
Estimate "Simple" SDE in Local Time Window. Then Apply Hypothesis Tests To Quantify Time of Validity



$$d\Phi_t = (A + B\Phi_t)dt + (C + D\Phi_t)dW_t$$

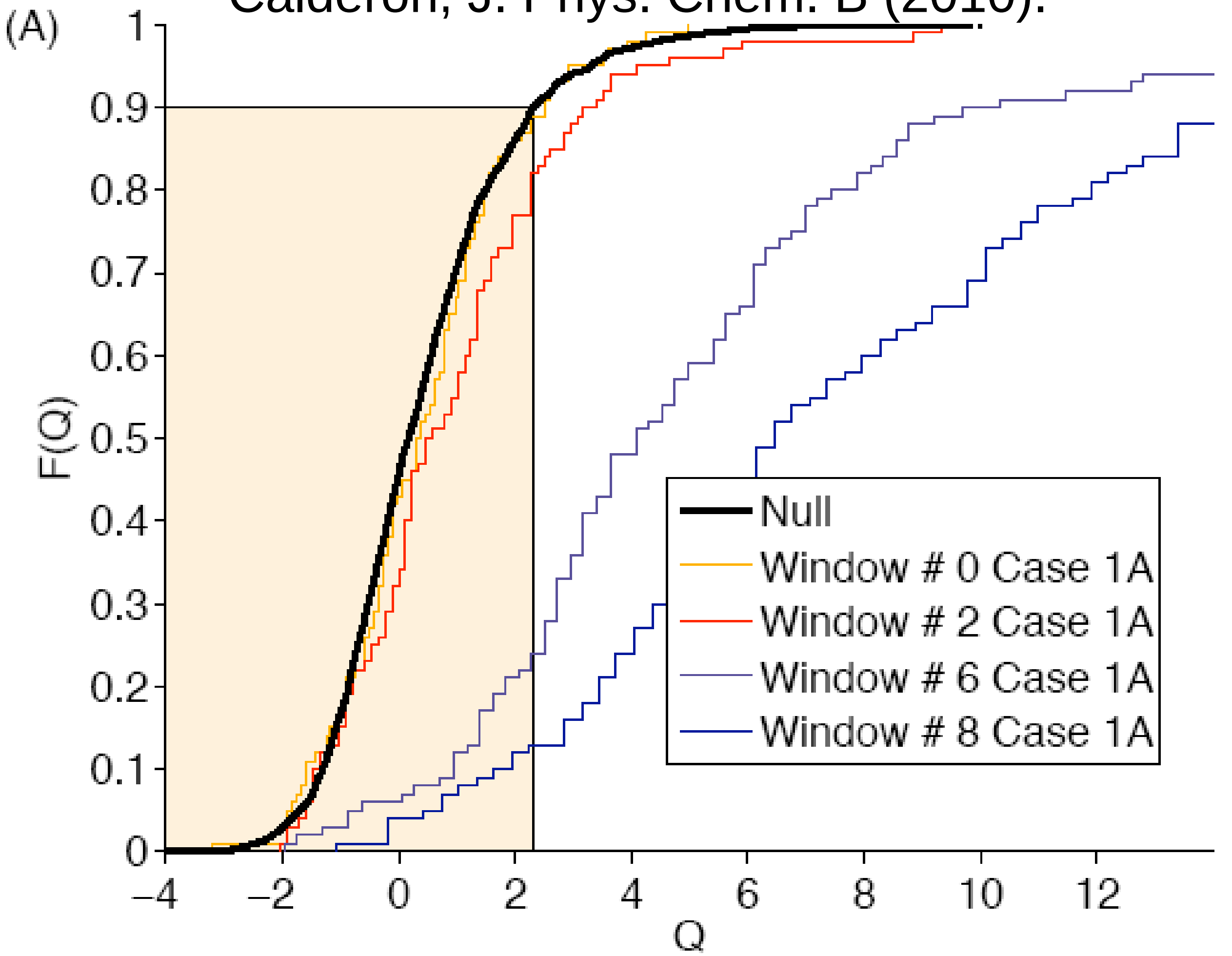
Estimate Model in Window "0"; Keep Fixed and Apply to Other Windows




Window 0
Window 2
Window 6
Window 8
−0.8
−0.9
−1
−1.1
−1.2
−1.3
−1.4
−1.5
−1.6
−1.7
y
0
0.2
0.4
0.6
0.8
1
Time [ns]

Each Test Statistic Uses ONE Short Path

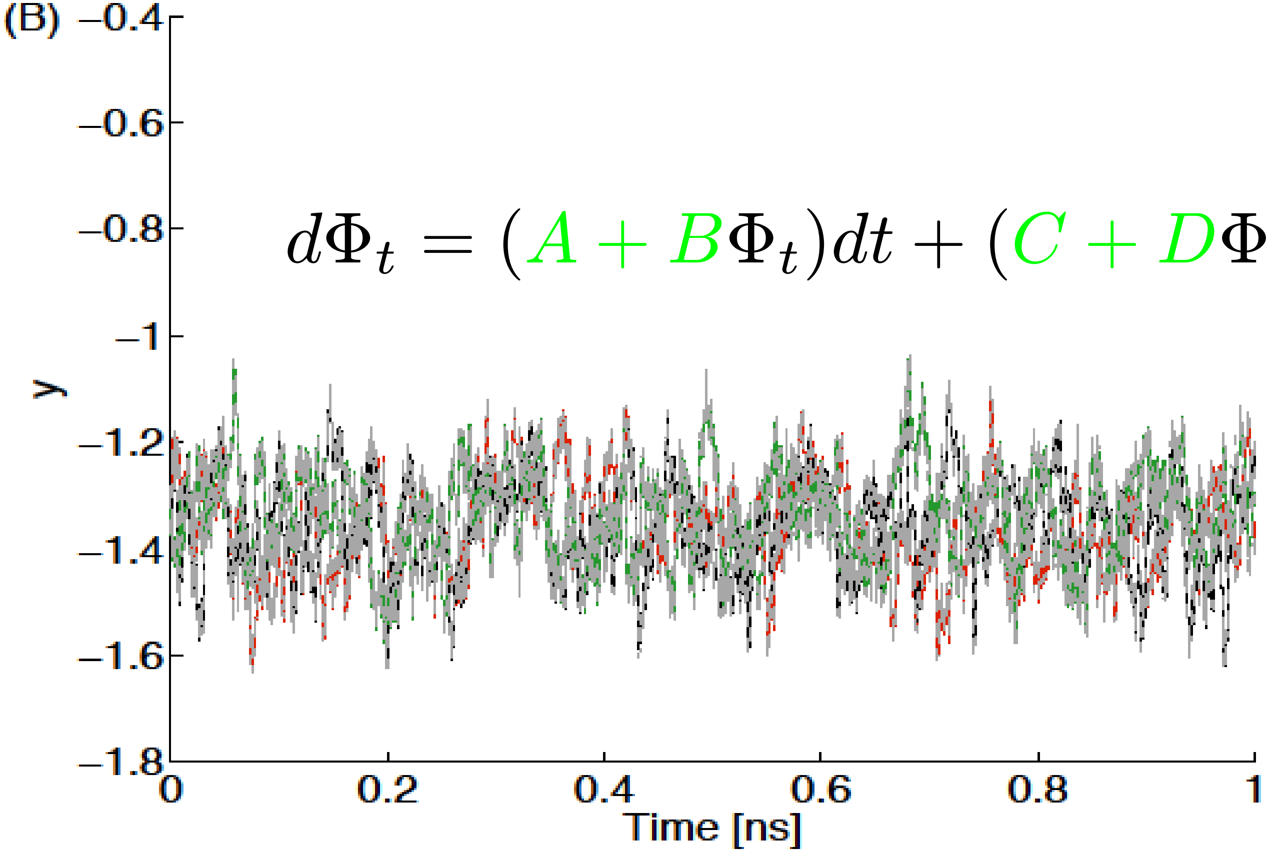
Calderon, J. Phys. Chem. B (2010).



Stationary Time Series?

Subtle Noise Magnitude Changes due Unresolved Degrees of Freedom (Check Validity of a "Born-Oppenheimer" Type Proxy)

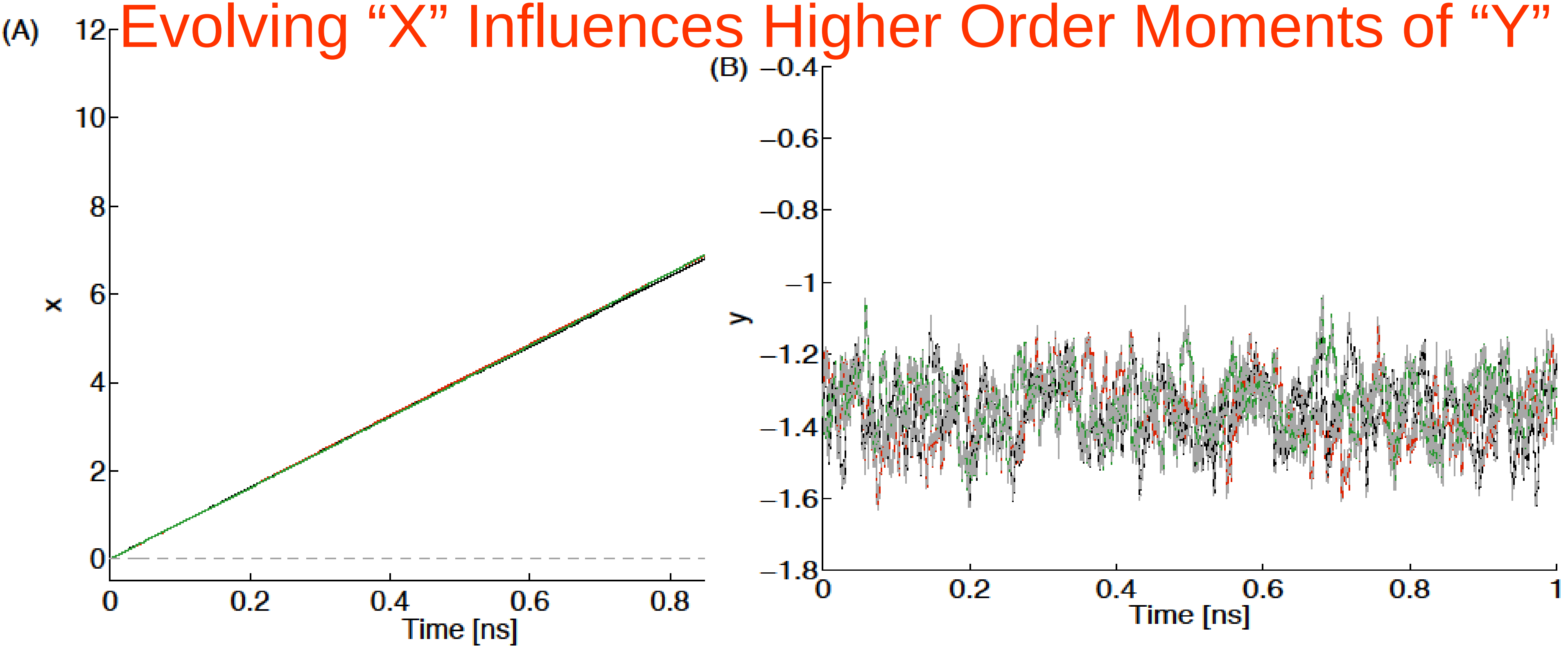
$$d\Phi_t = (A + B\Phi_t)dt + (C + D\Phi_t)dW_t$$



Stationary Time Series?

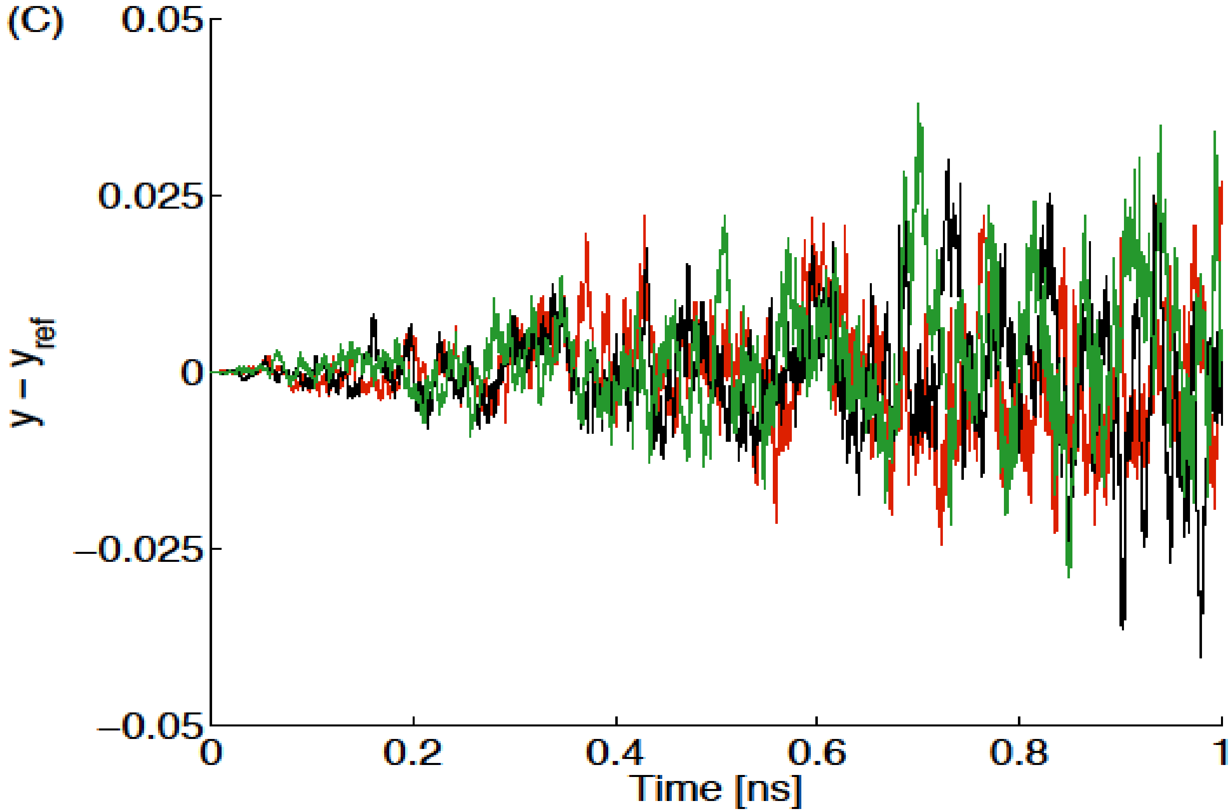
Subtle Noise Magnitude Changes due Unresolved Degrees of Freedom (Check Validity of a “Born-Oppenheimer” Type Proxy)

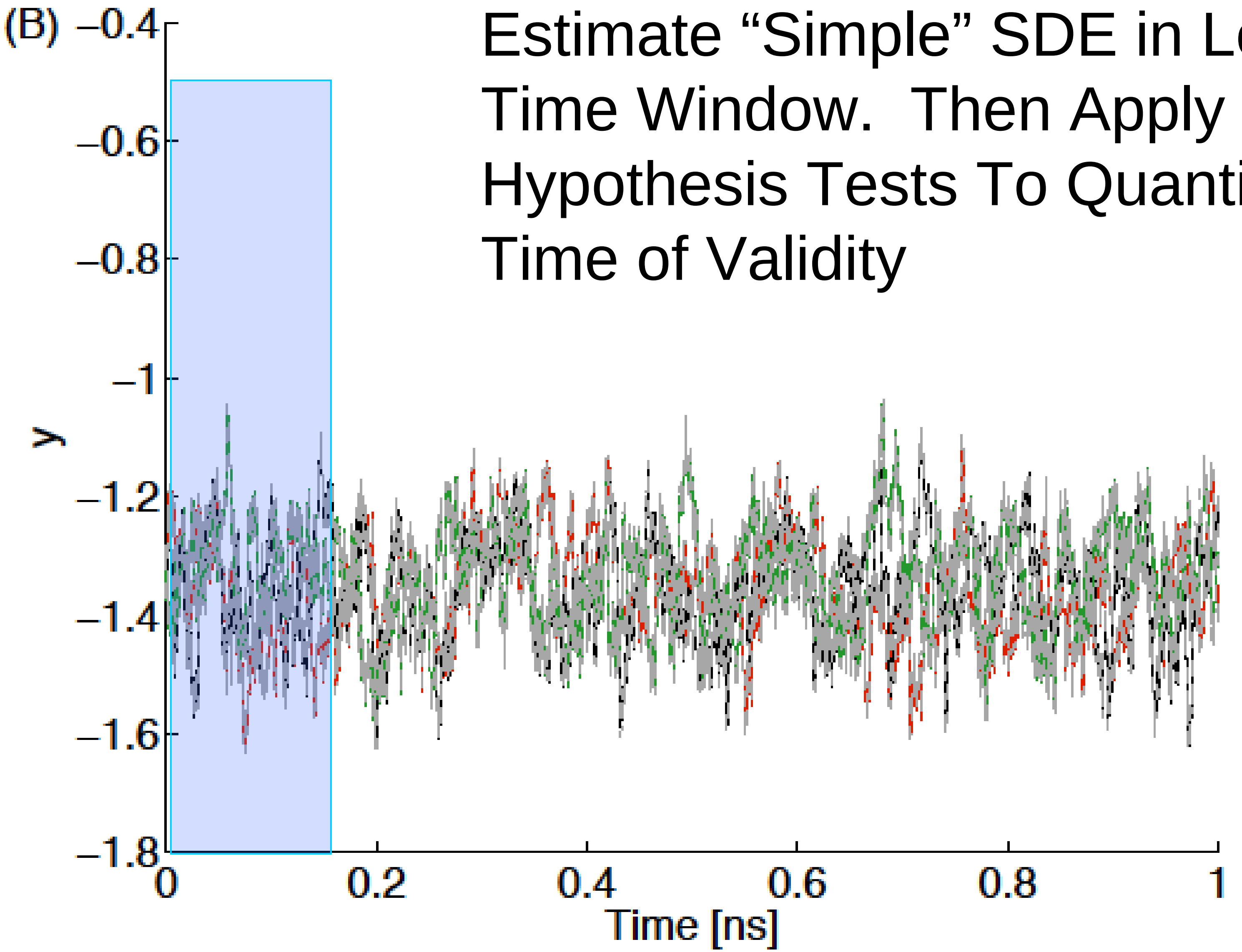
Evolving “X” Influences Higher Order Moments of “Y”



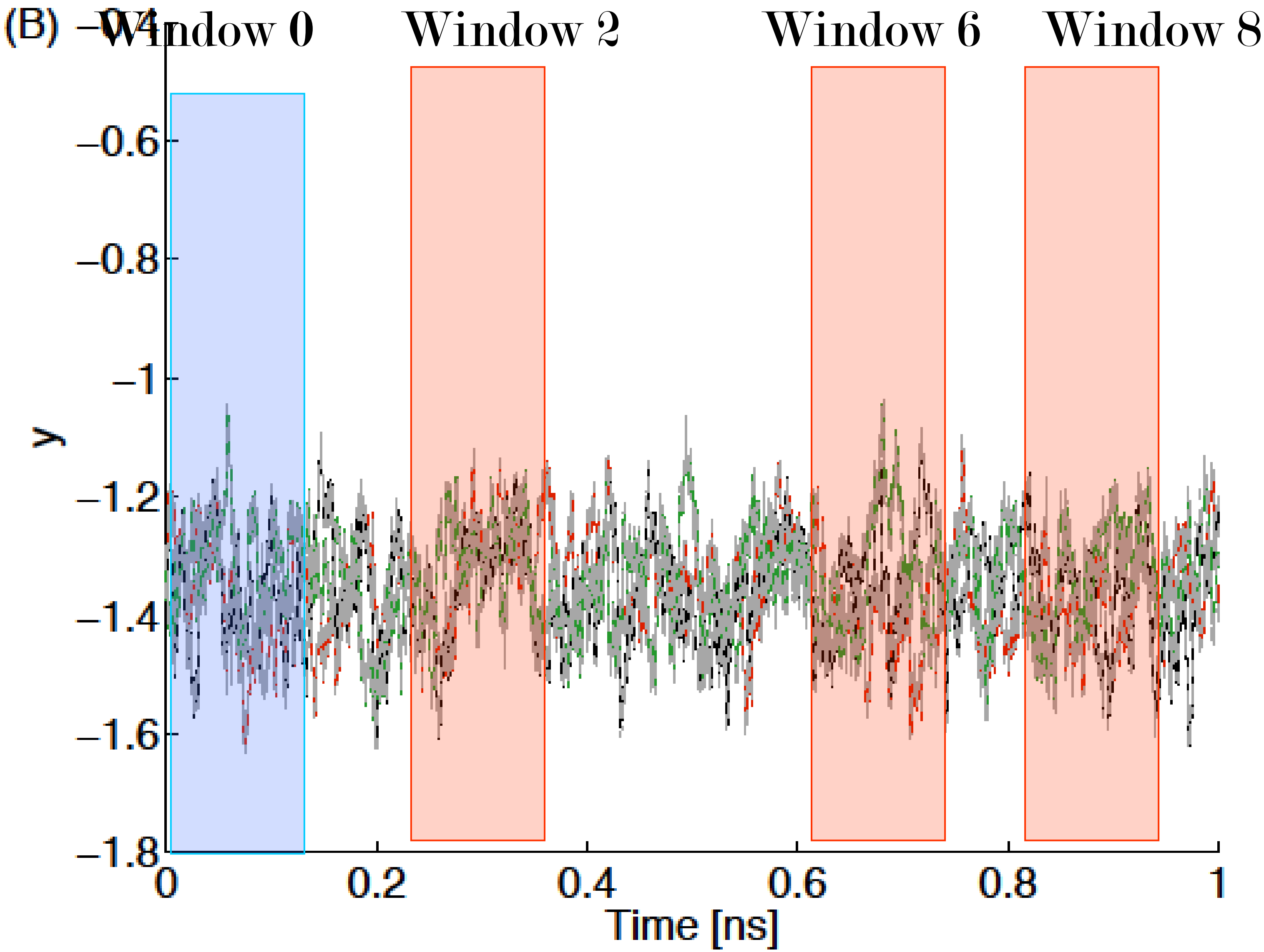
Stationary Time Series?

Subtle Noise Magnitude Changes due Unresolved Degrees of Freedom (Check Validity of a “Born-Oppenheimer” Type Proxy)



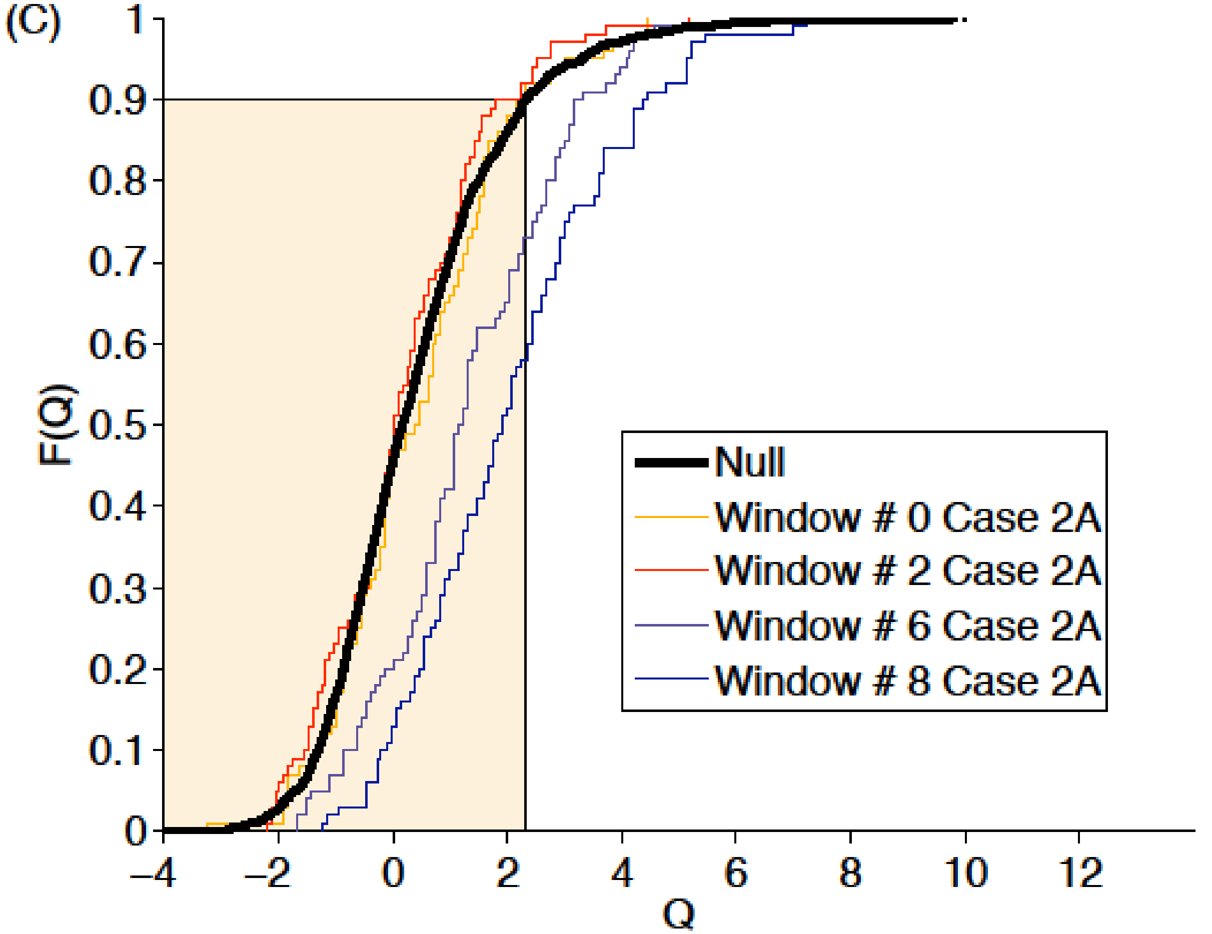


Estimate “Simple” SDE in Local Time Window. Then Apply Hypothesis Tests To Quantify Time of Validity


Window 0
Window 2
Window 6
Window 8
(B)
−0.4
−0.6
−0.8
−1
−1.2
−1.4
−1.6
−1.8
y
0
0.2
0.4
0.6
0.8
1
Time [ns]

"Subtle" Noise Magnitude Changes

Calderon, J. Phys. Chem. B (2010).



Frequentist vs. Bayesian Approaches

- Concept of “Residual” Unnatural in Bayesian Setting. All Information Condensed into Posterior (Time Residual Can Help in Non-ergodic Settings)

- Uncertainty Information Available in Both Under Ideal Models with Heavy Computation (Bayesian Methods Need “Prior” Specification)

- Test Against Many Alternatives in a Frequentist Setting (Instead of a Handful of Candidate Models as is Done in Bayesian Methods)

Several Issues

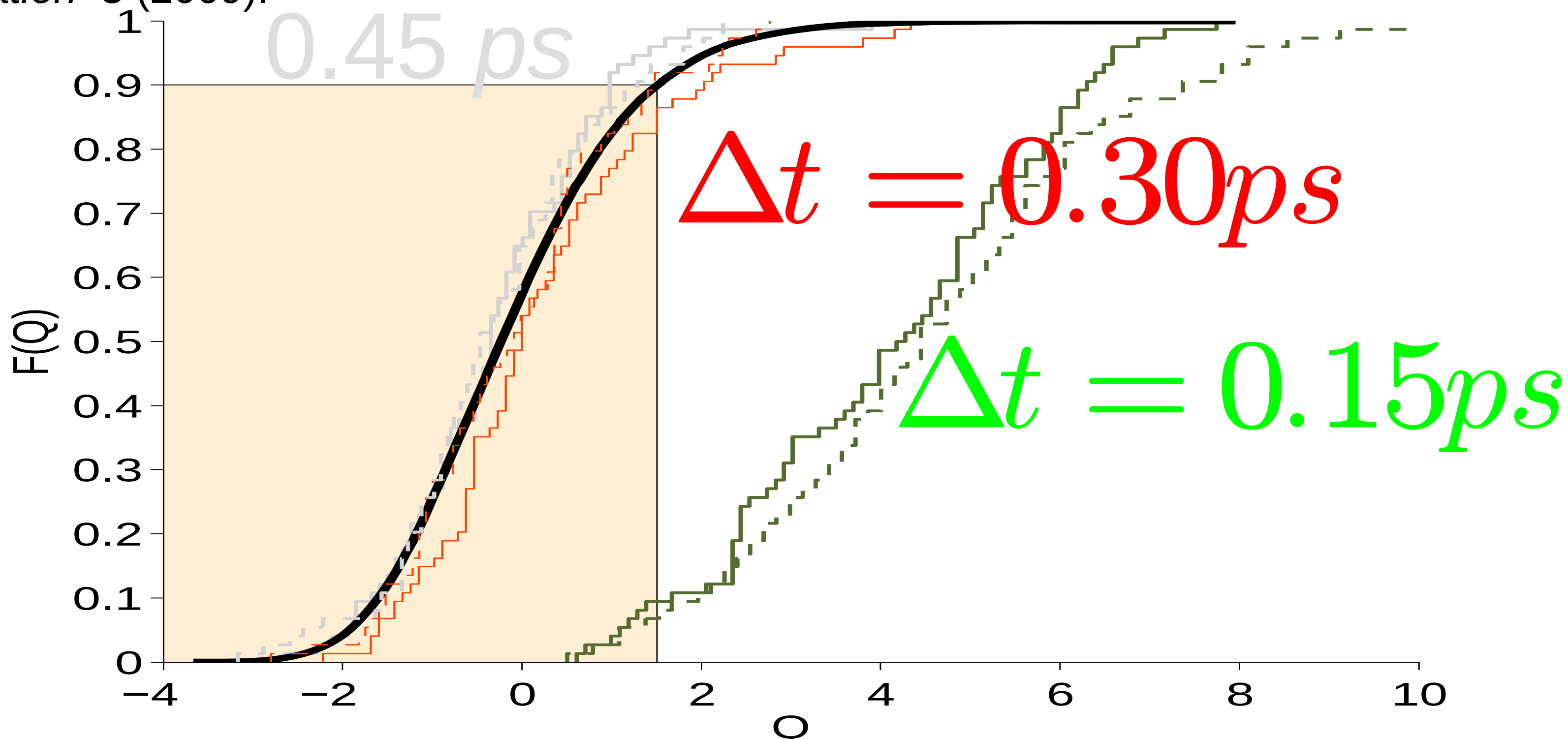
Evaluate Z at Estimated Parameter (Their Limit Distribution Requires Root N Consistent Estimator to Get [Asymptotic] Critical Values)

Asymptopia is a Strange Place [Multiplying by Infinity Can Be Dangerous When Numerical Proxies are Involved]

In Nonstationary Case, Initial Condition Distribution May Be Important but Often Unknown.

Testing Surrogate Models

Hong & Li, *Rev. Financial Studies*, **18** (2005). Chen, S, Leung, D. & Qin,J., *Annals of Statistics*, **36** (2008). Ait-Sahalia, Fan, Peng, *JASA* (in press) Calderon & Arora, *J. Chemical Theory & Computation* **5** (2009).



One Times Series Reduces to One Test Statistic. Empirical CDF Summarizes Results from 100 Time Series

A Sketch of Local to Global

or

Fitting Nonlinear SDEs Without A Priori Knolwedge of Global Parametric Model

Local Time Dependent Diffusion Models

Calderon, *J. Chem. Phys.* **126** (2007).

$$dz_t = b(z_t, t; \Theta)dt + \sigma(z_t; \Theta)dW_t$$

Approximate Locally by Low Order Polynomials, e.g.

$$b(z) \approx A + Bz + f(t)$$

$$\sigma(z) \approx C + Dz$$

Can use Physics-based Proxies e.g. Overdamped Langevin Dynamics

$$\sigma(z) \approx C + Dz$$

$$b(z) \approx (C + Dz)^2/(2k_BT)(A + Bz + f(t))$$

Maximum Likelihood

For given model & discrete observations, find the maximum likelihood estimate:

$$\hat{\theta} \equiv \max_{\theta}\ p(z_0, \ldots, z_T;\ \theta)$$

Special case of Markovian Dynamics:

$$\hat{\theta} \equiv \max_{\theta}\ p(z_0; \Theta)p(z_1|z_0; \theta) \ldots p(z_T|z_{T-1}; \theta)$$

“Transition Density” (aka Conditional Probability Density)

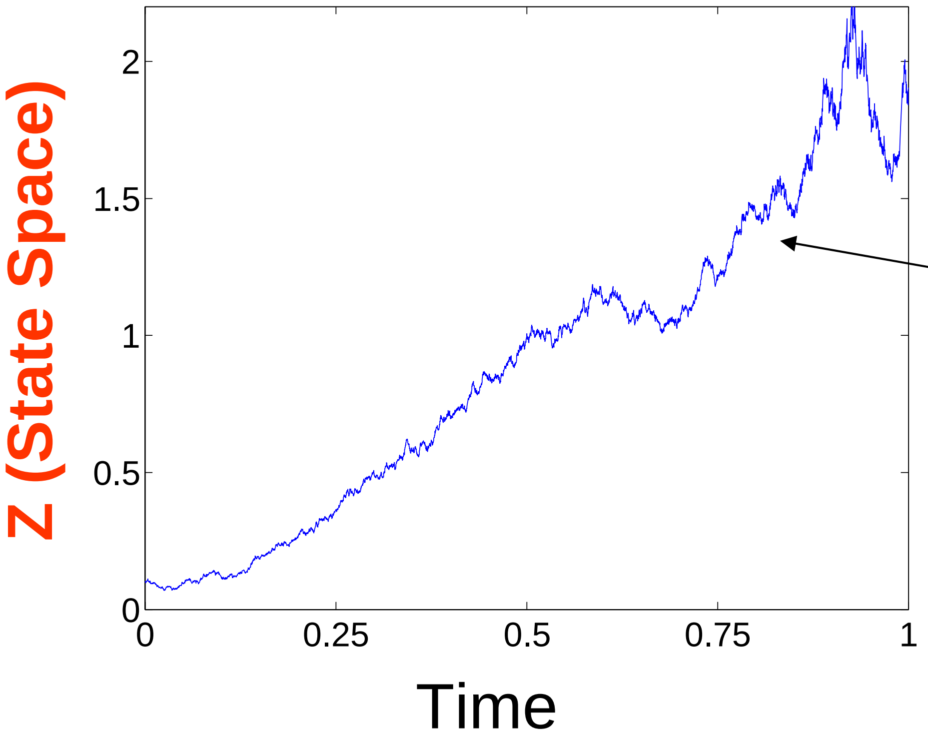
After Estimating Acceptable (Not Rejected) Local Models, What's Next?

- Quality of Data: "Variance" (MC Simulation in Local Windows)

- Then Global Fit (Nonlinear Regression)

Note: Noise Magnitude Depends on State

And the State is Evolving in a Non-Stationary Fashion

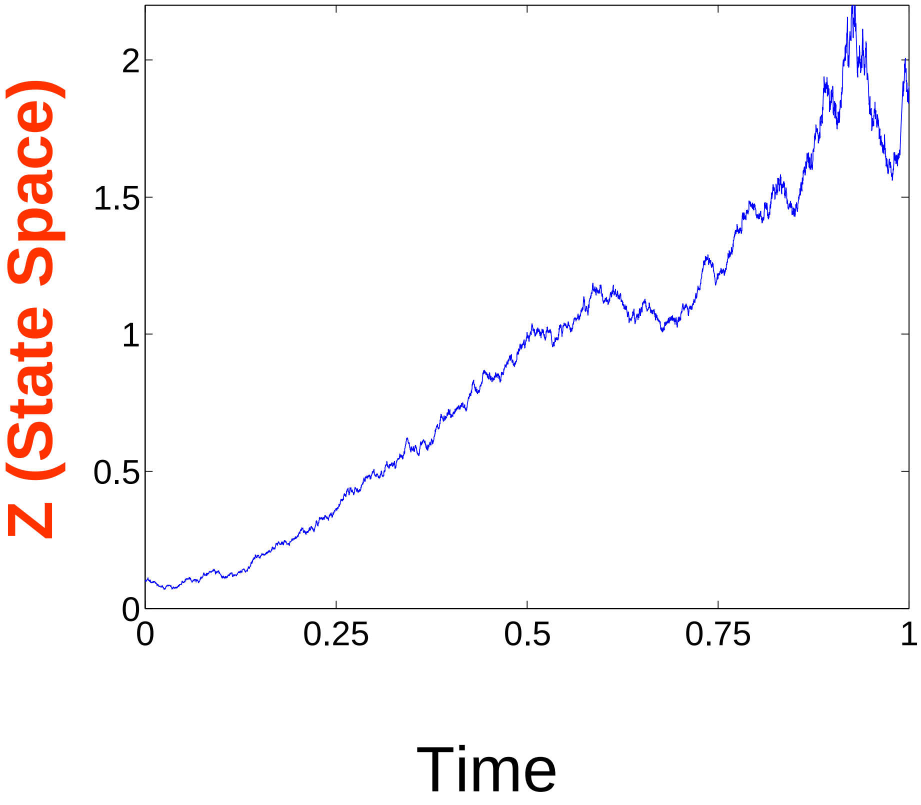


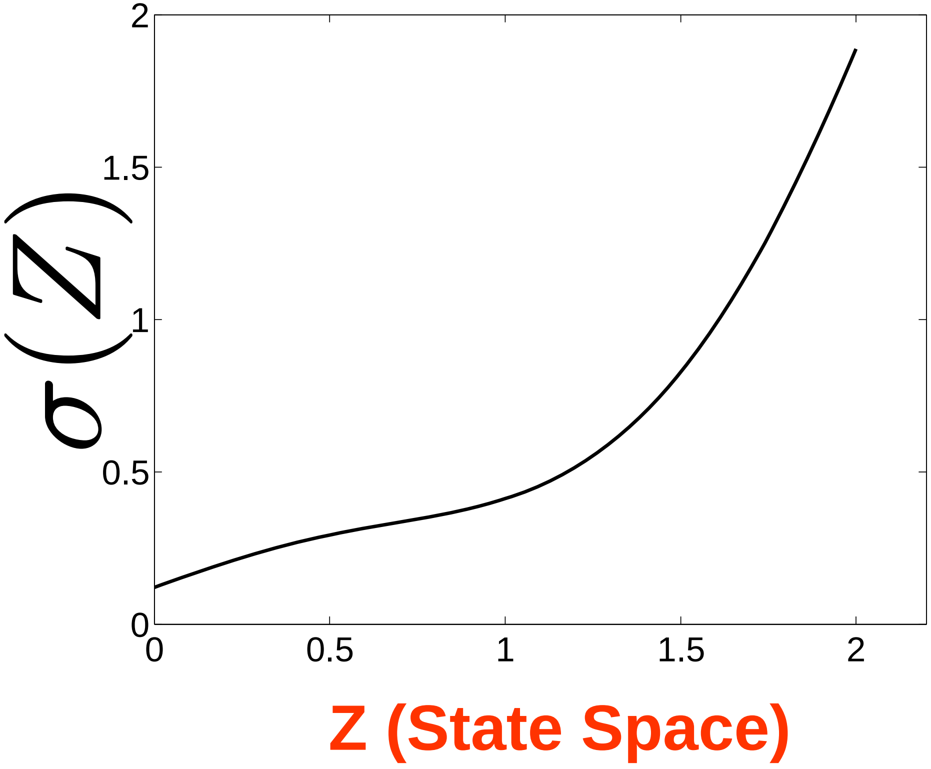
State Dependent Noise Would "Fool" Wavelet Type Methods

$$dZ_t = k(v_{\mathrm{pull}}t - Z_t)dt + \sigma(Z_t)dW_t$$

Note: Noise Magnitude Depends on State

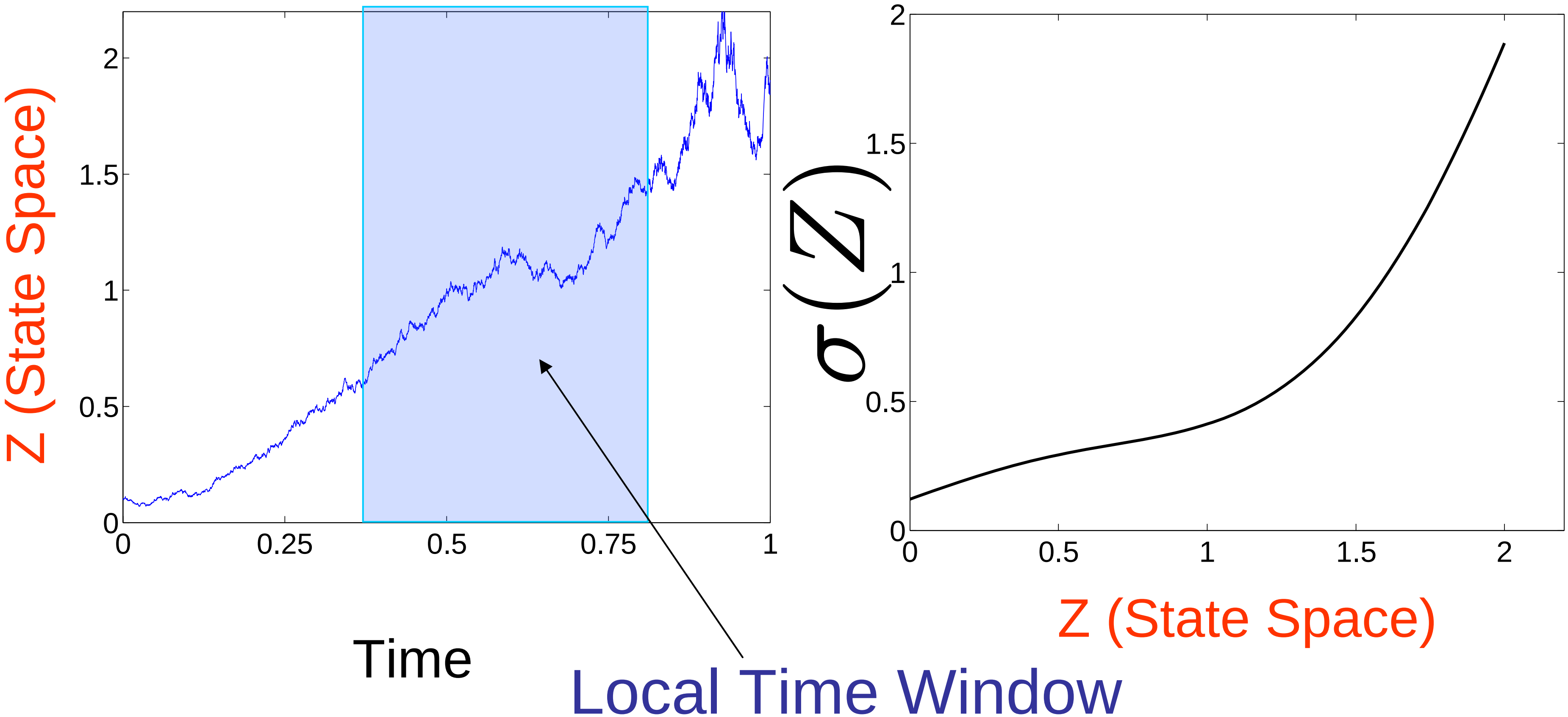
And the State is Evolving in a Non-Stationary Fashion

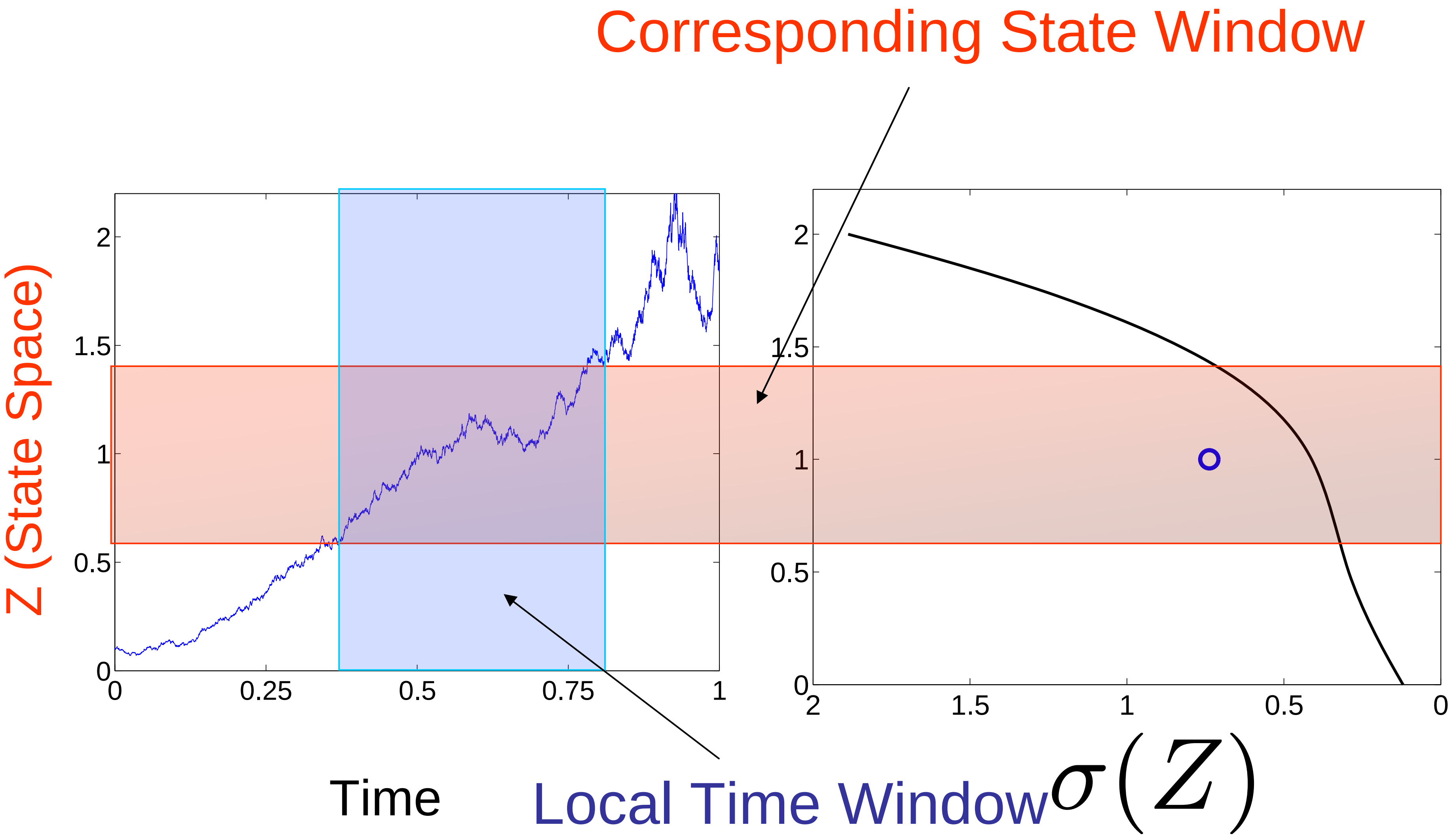




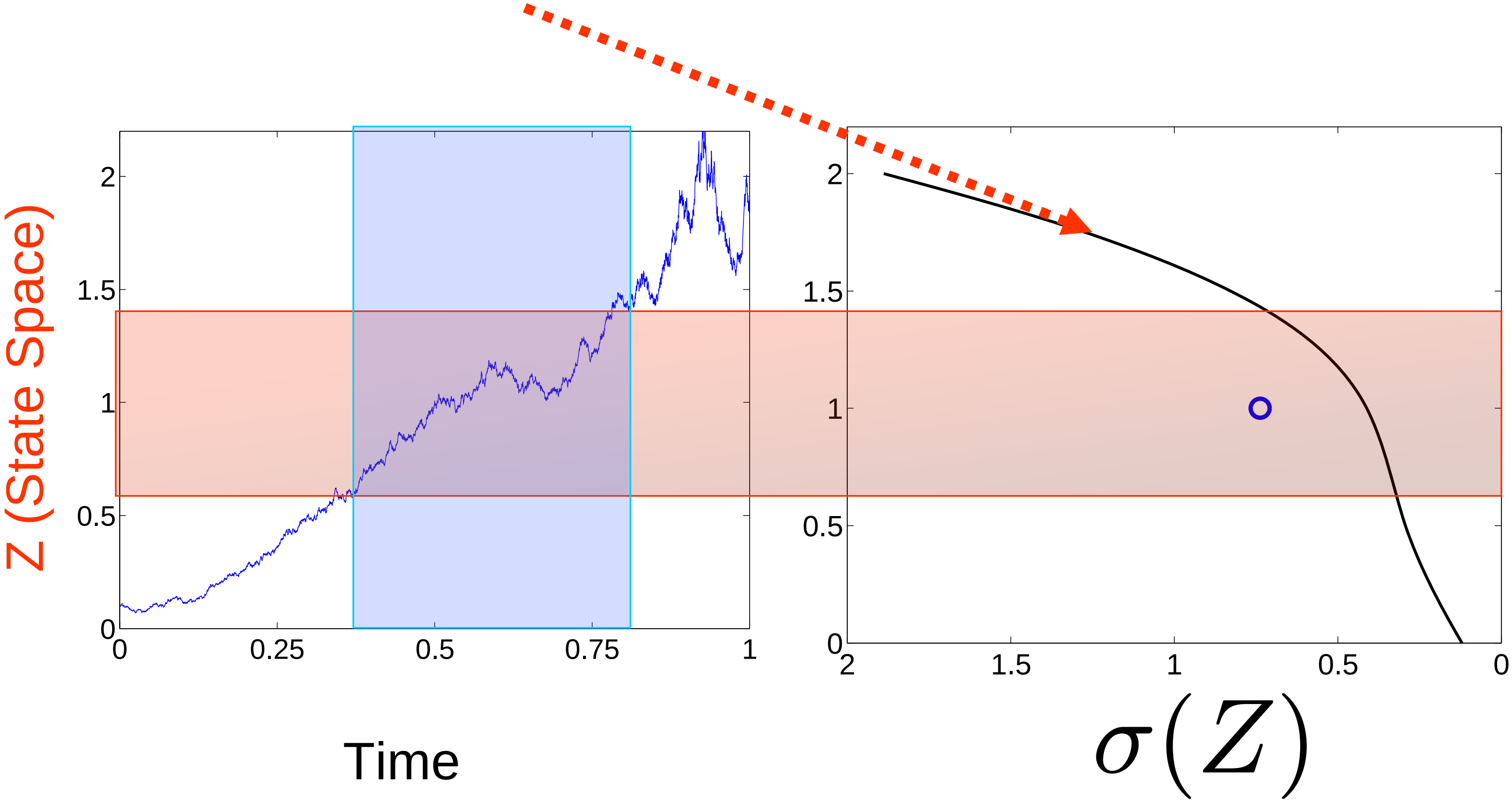
Note: Noise Magnitude Depends on State

And the State is Evolving in a Non-Stationary Fashion

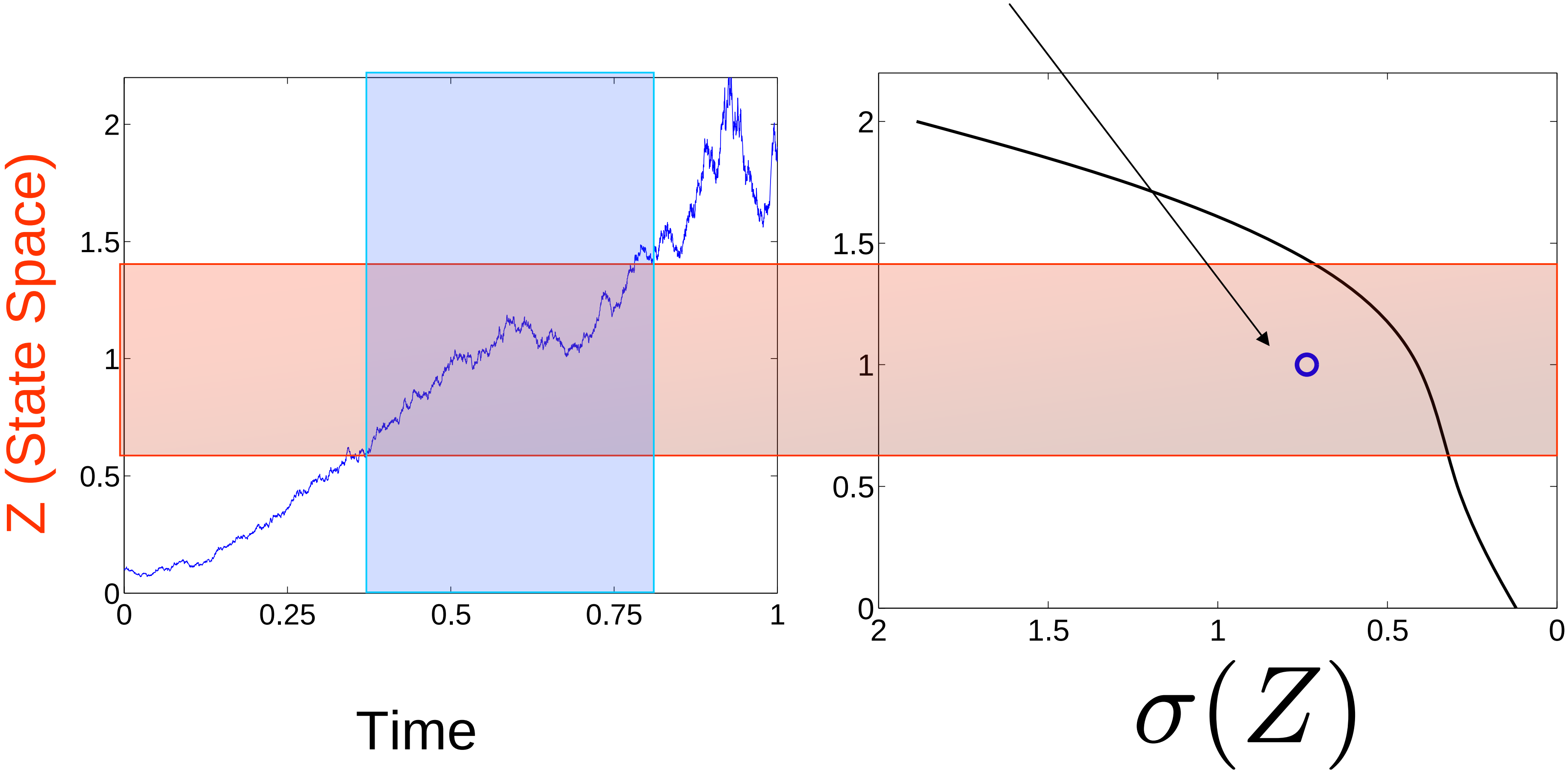



Corresponding State Window
Z (State Space)
2
1.5
1
0.5
0
0
0.25
0.5
0.75
1
Time
Local Time Window
2
1.5
1
0.5
0
2
1.5
1
0.5
0
$\sigma(Z)$

Global Nonlinear Function of Interest (the “truth”)

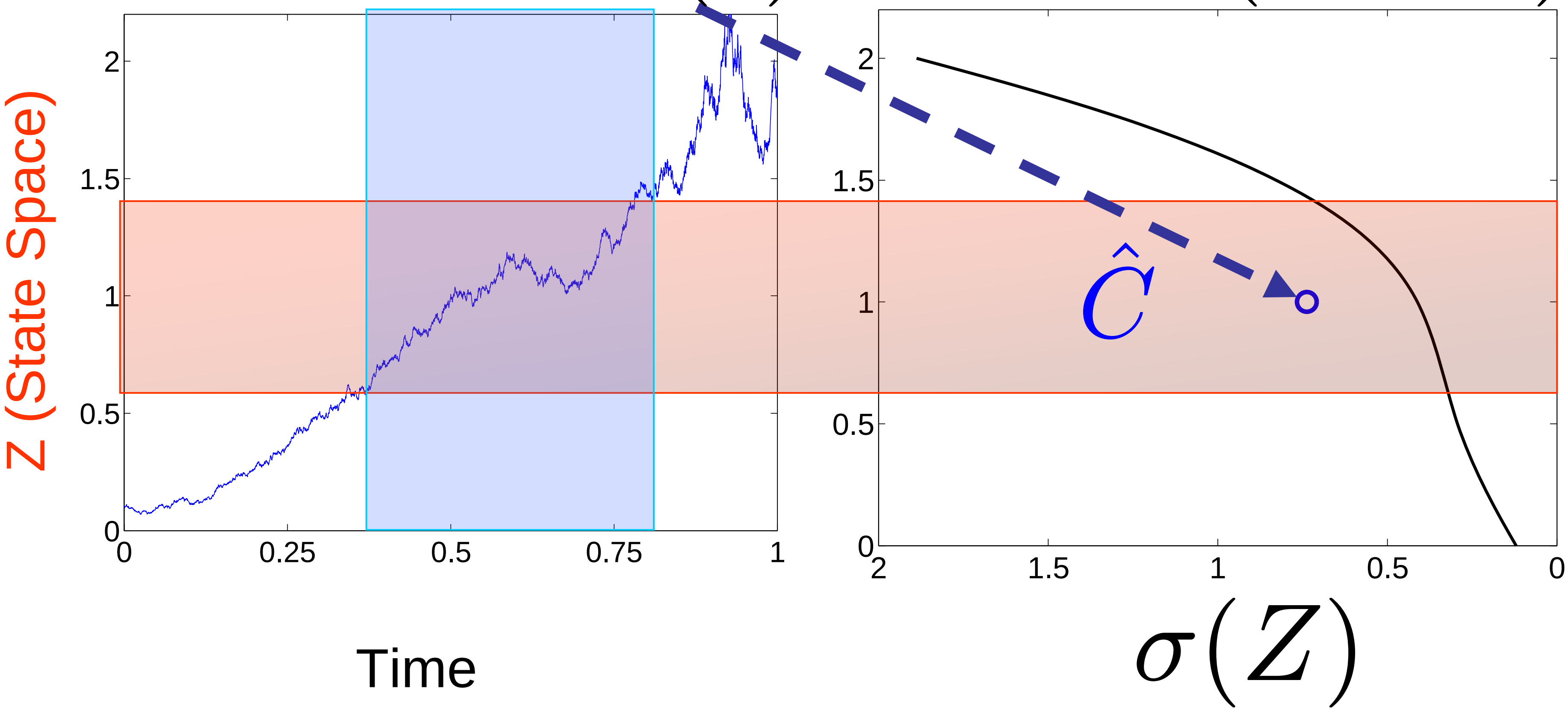


Point Estimate Comes From Local Parametric Model (Parametric Likelihood Inference Tools Available)



Point Estimate Comes From Local Parametric Model (Parametric Likelihood Inference Tools Available)

$$\sigma(z) \approx C + D(z - z^0)$$



Maximum Likelihood

For given model & discrete observations, find the maximum likelihood estimate:

$$\hat{\theta} \equiv \max_{\theta} \; p(z_0, \ldots, z_T; \; \theta)$$

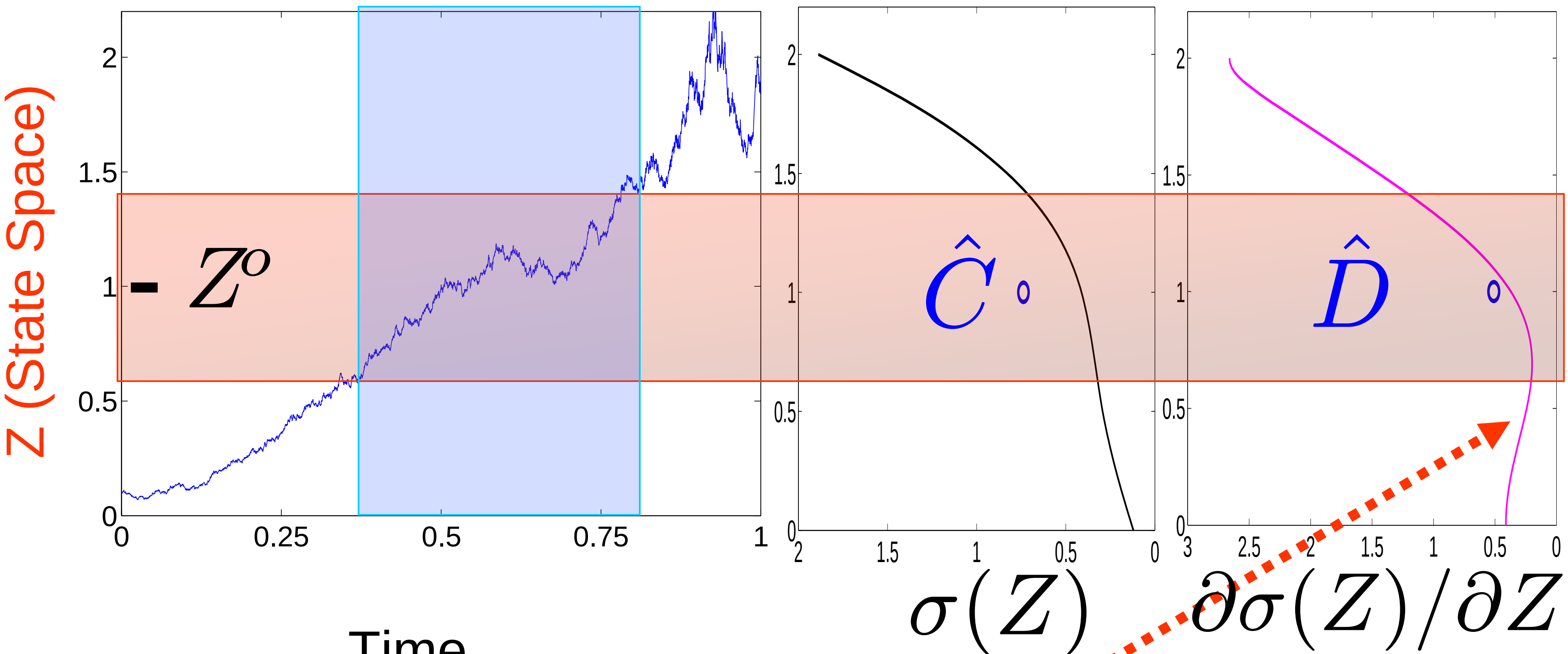
Special case of Markovian Dynamics:

$$\hat{\theta} \equiv \max_{\theta} \; p(z_0; \Theta) p(z_1 | z_0; \theta) \ldots p(z_T | z_{T-1}; \theta)$$

"Transition Density" (aka Conditional Probability Density)

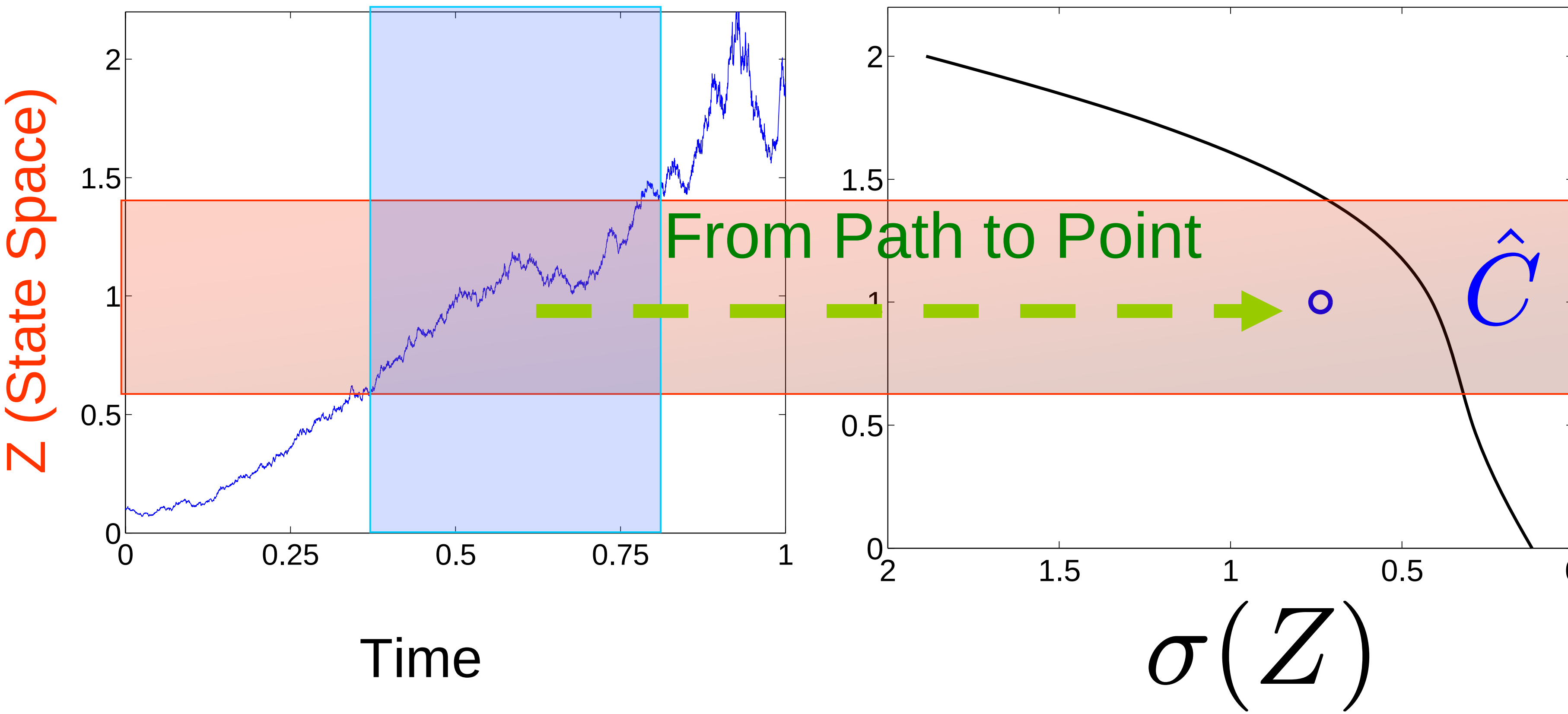
Noisy Point Estimates (finite discrete time series sample uncertainty)

$$\sigma(z) \approx C + D(z - z^0)$$

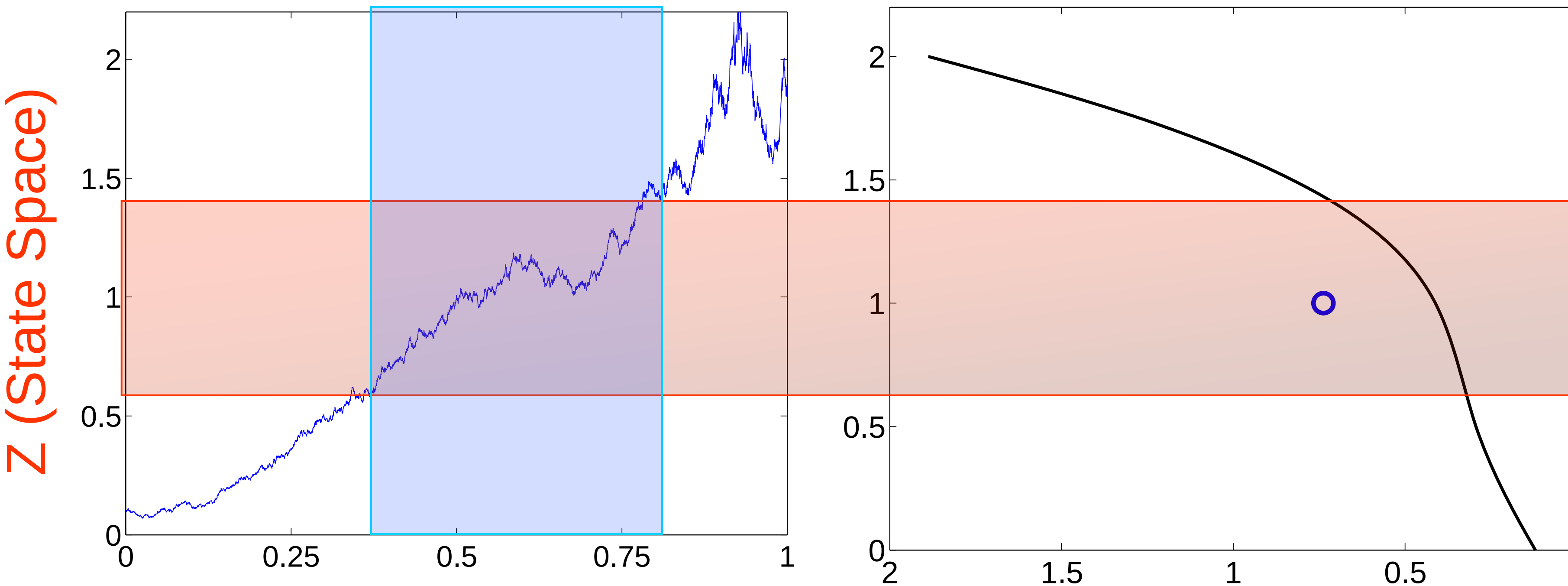


Spatial Derivative of Function of Interest

Idea Similar in spirit to J. Fan, Y. Fan, J. Jiang, *JASA* **102** (2007) except uses fully parametric MLE (no local kernels) in discrete local state windows

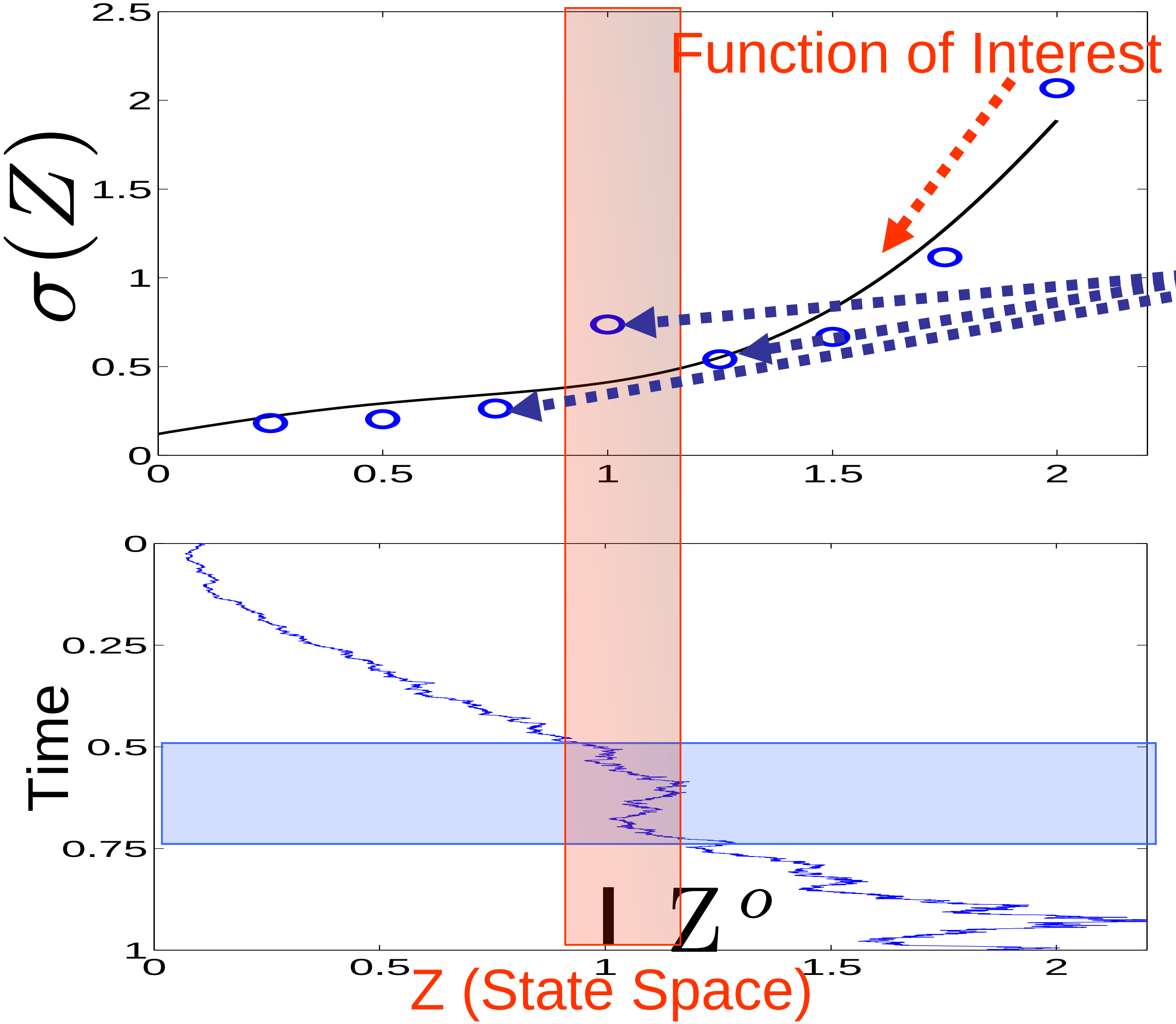


HOWEVER: Not Willing To Assume Stationarity in Windows. (Completely **Nonparametric** and Fourier Analysis **Problematic**). "Subject Specific Function Variability" of Interest



Validity of Local Models Can Be **Tested** Using Generalized Residuals Available Using Transition Density with Local Parametric Approach

$\sigma(Z)$

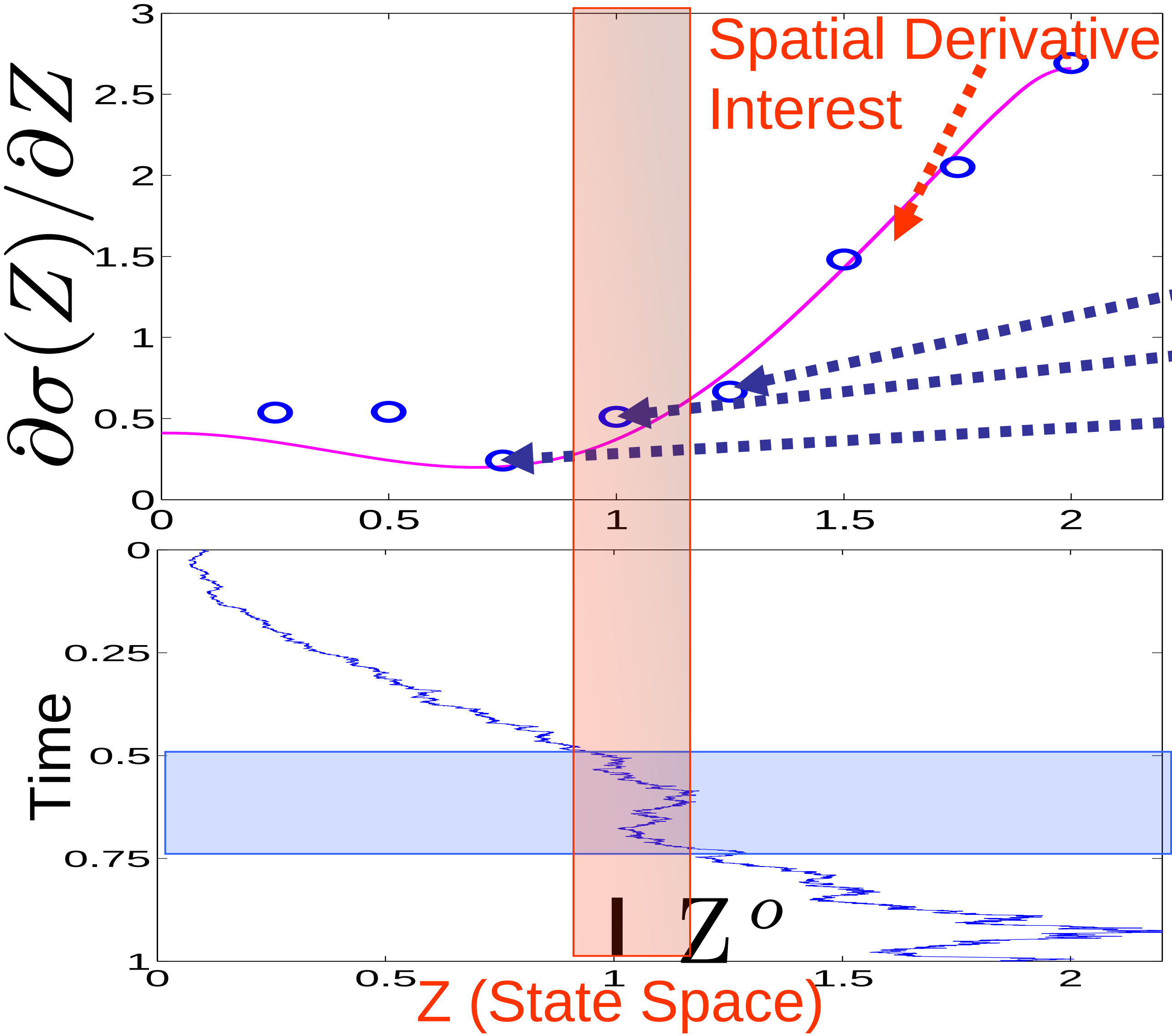


Noisy Point Estimates

(Obtained Using Local Maximum Likelihood)

Transform "X vs t" Data to "X vs Y=F(X)"

Estimate Nonlinear Function via Regression

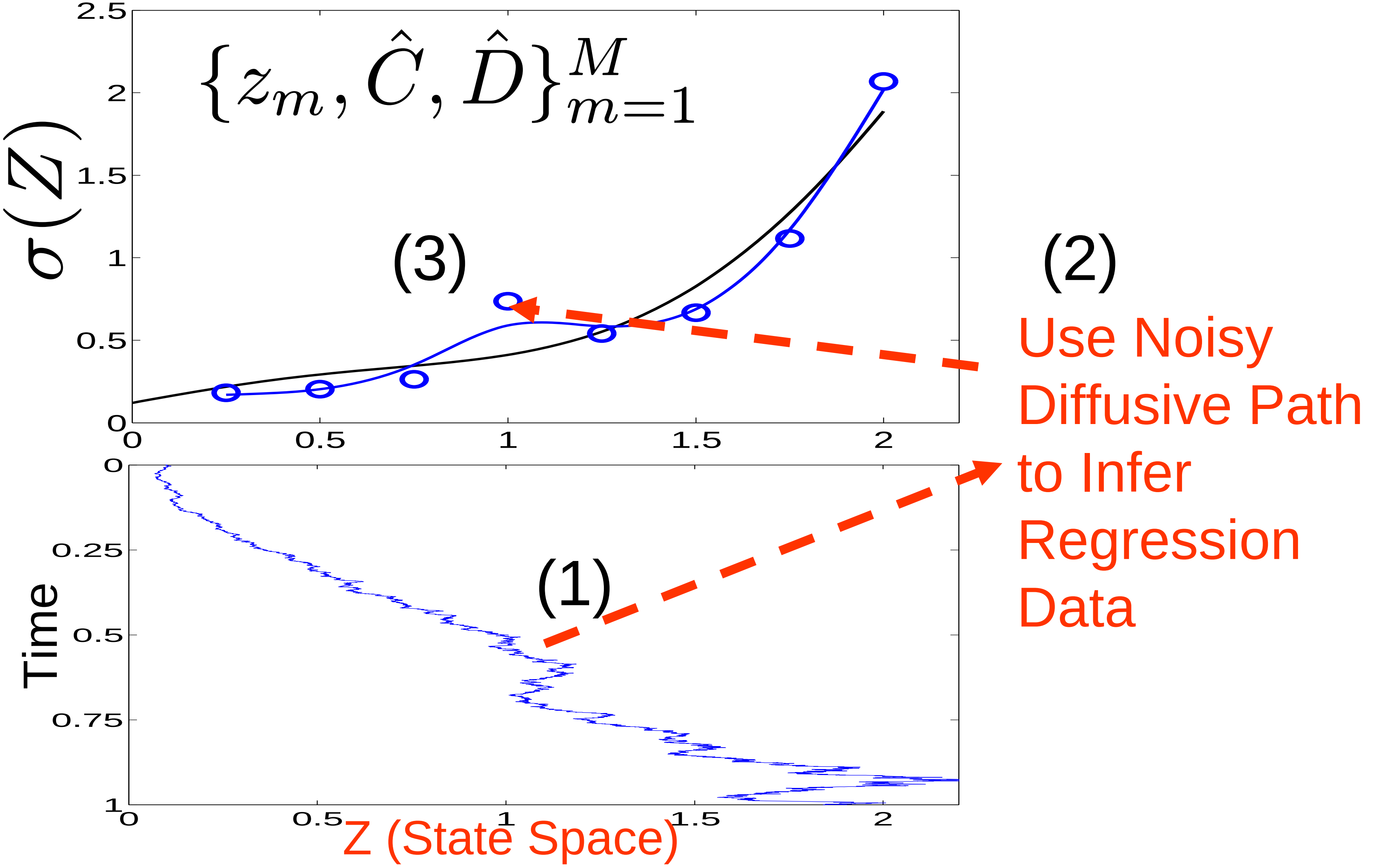


Noisy Point Estimates

(Obtained Using Local Maximum Likelihood)

Transform "X vs t" Data to "X vs Y'=G(X)"

Estimate Nonlinear Function via Regression

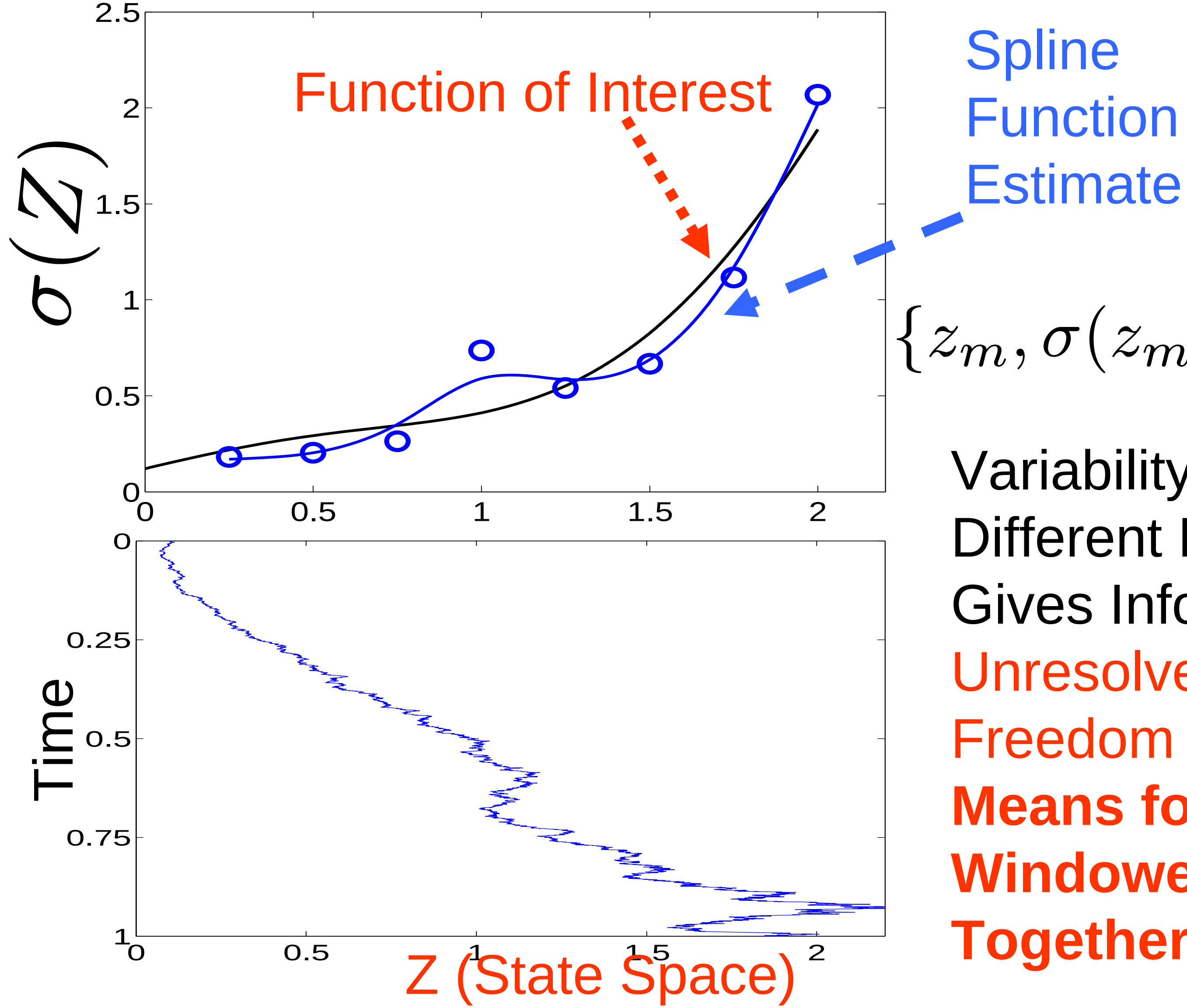

$\{z_m, \hat{C}, \hat{D}\}_{m=1}^{M}$
(3)
(2)
Use Noisy Diffusive Path to Infer Regression Data
(1)
$\sigma(Z)$
Time
Z (State Space)
2.5
2
1.5
1
0.5
0
0.25
0.75

Same Ideas Apply to Drift Function:

$$\{z_m, \hat{A}, \hat{B}\}_{m=1}^{M}$$

Can Also Entertain Simultaneous Regression:

$$\{z_m, \hat{A}, \hat{B}, \hat{C}, \hat{D}\}_{m=1}^{M}$$



$$\{z_m, \sigma(z_m), \frac{d\sigma(z)}{dz}\Big|_{z_m}\}_{m=1}^{M}$$

Variability of Between Different Functions Gives Information About Unresolved Degrees of Freedom **Reliable Means for Patching Windowed Estimates Together is Desirable**

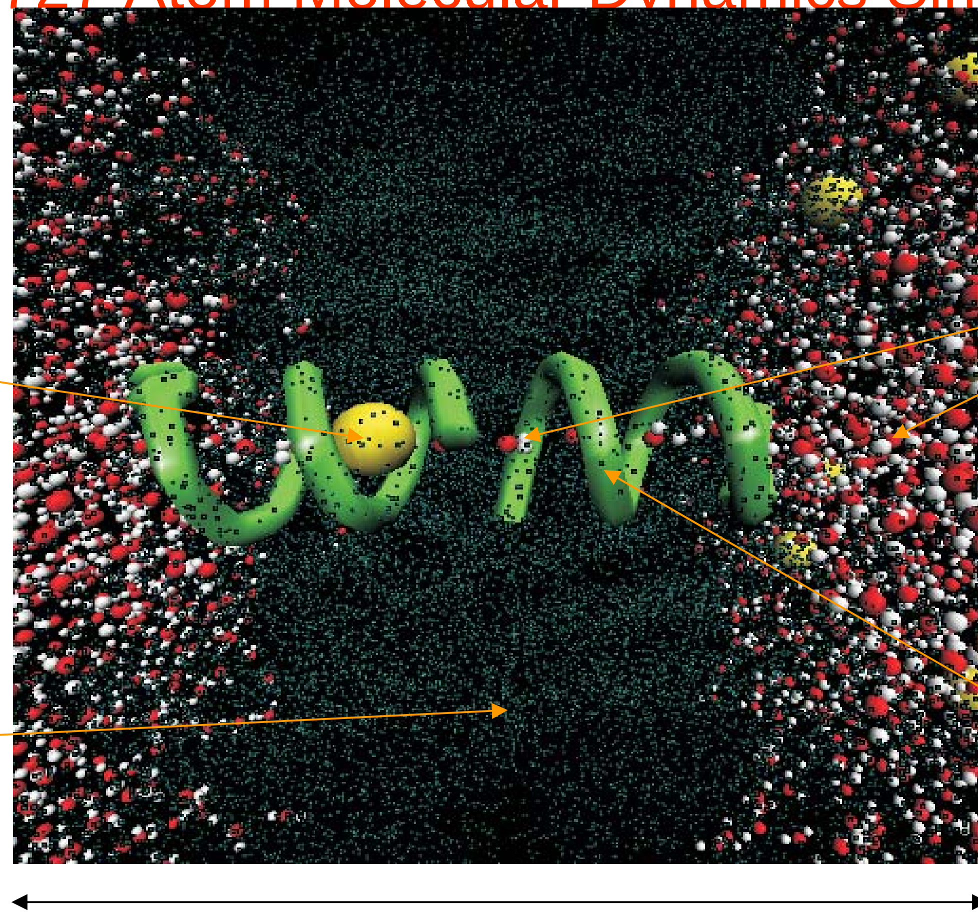
Open Mathematical Questions for Local SDE Inference

How Can One Efficiently Choose "Optimal" Local Window Sizes" ?

Hypothesis Tests Checking Markovian Assumption with Non-Stationary Data? ("Omnibus" goodness-of-fit tests of Hong and Li [2005] based on probability integral transform currently employed)

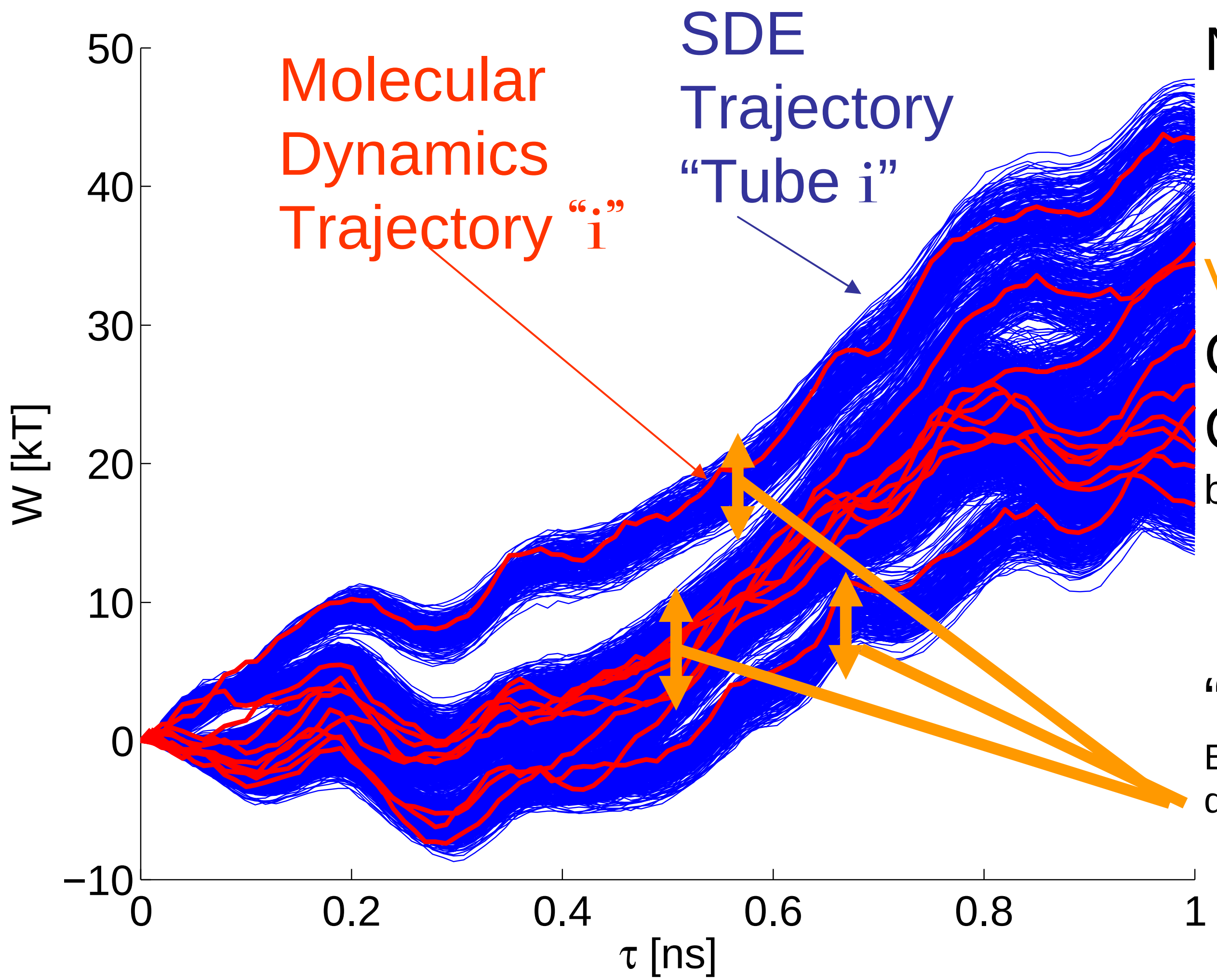
Gramicidin A: Commonly Studied Benchmark

36,727 Atom Molecular Dynamics Simulation



Compute Work: Integral of “Force time Distance”

PMF and Non-equilibrium Work



Nonergodic Sampling:

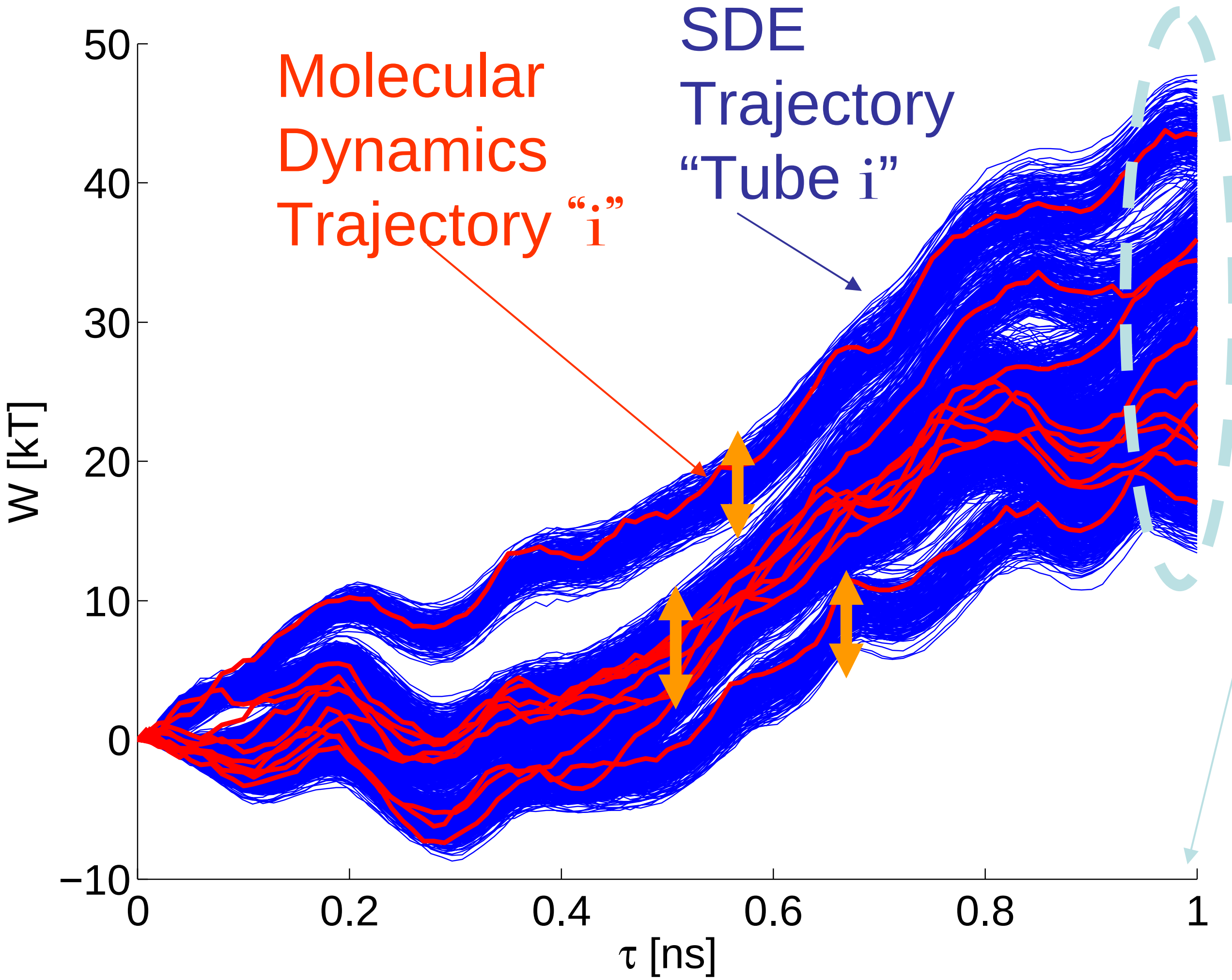
Pulling "Too Fast"

Variability Induced by Conformation Initial Condition (variability between curves)

and

"Thermal" Noise (Solvent Bombardment, Vibrational Motion, etc. quantified by tube width)

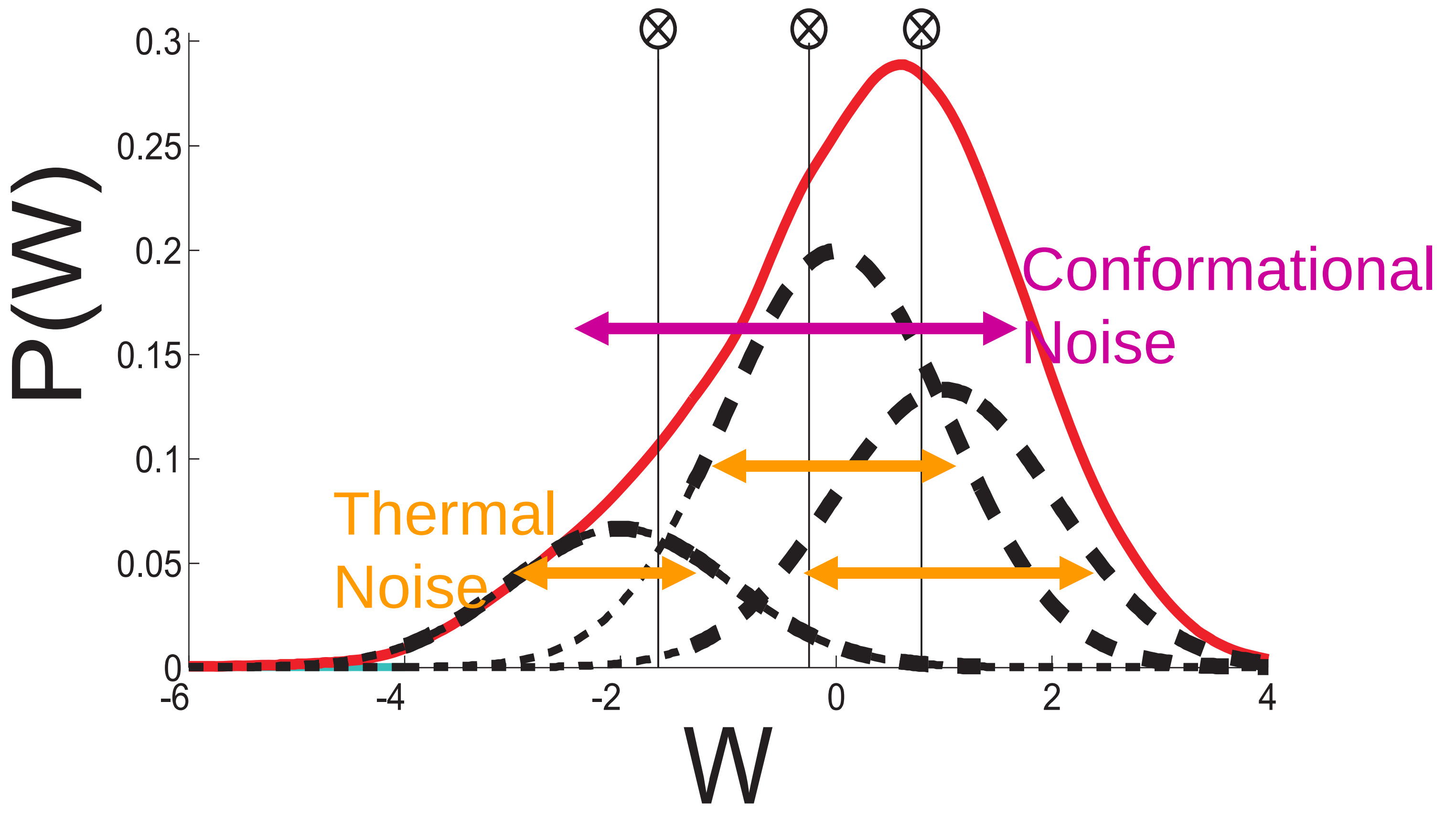
PMF and Non-equilibrium Work



Distribution at Final Time is non-Gaussian and can be viewed as MIXTURE of distributions

Importance of Tube Variability

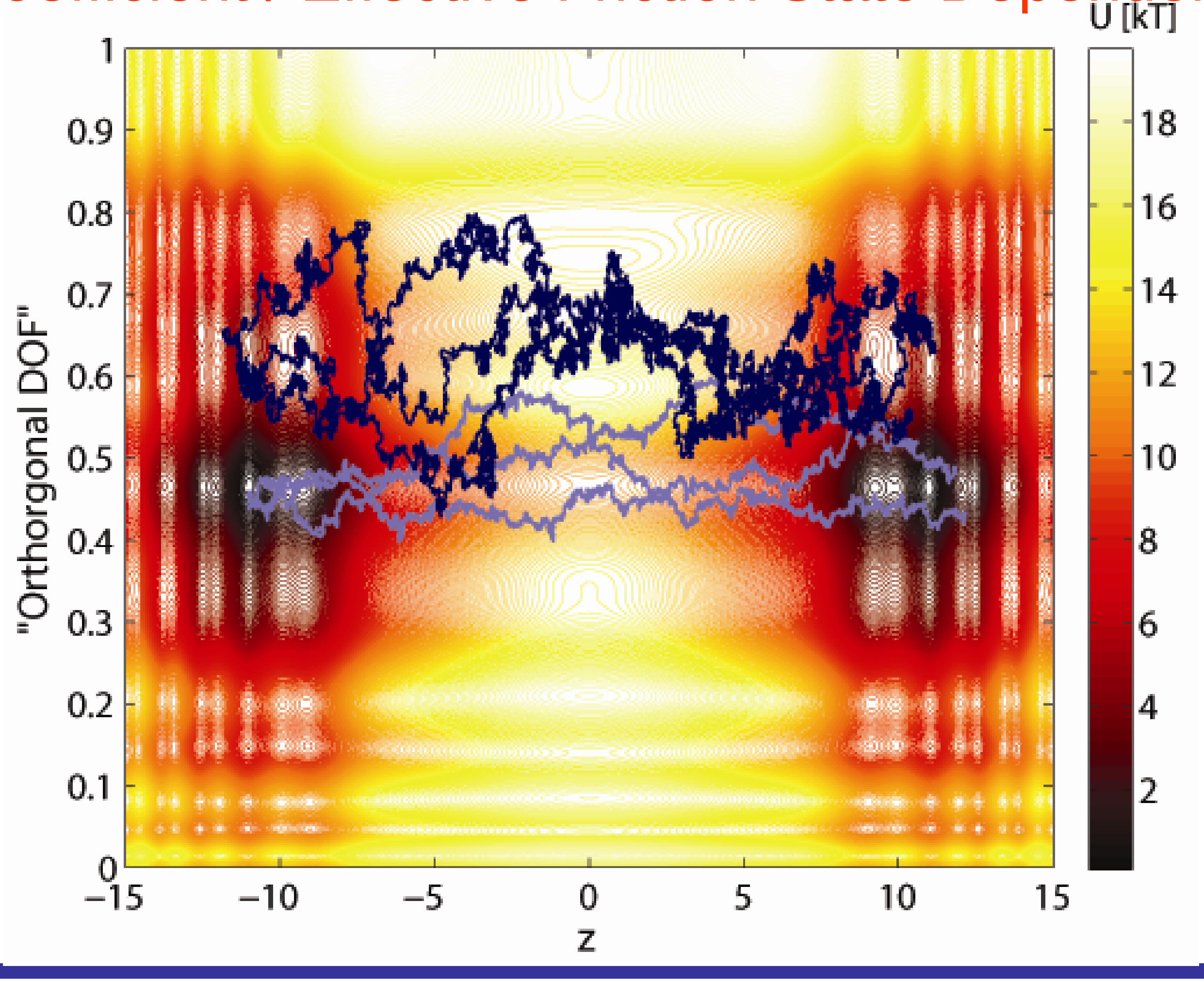
Mixture of Distributions



Other Kinetic Issues

Diffusion Coefficient / Effective Friction State Dependent

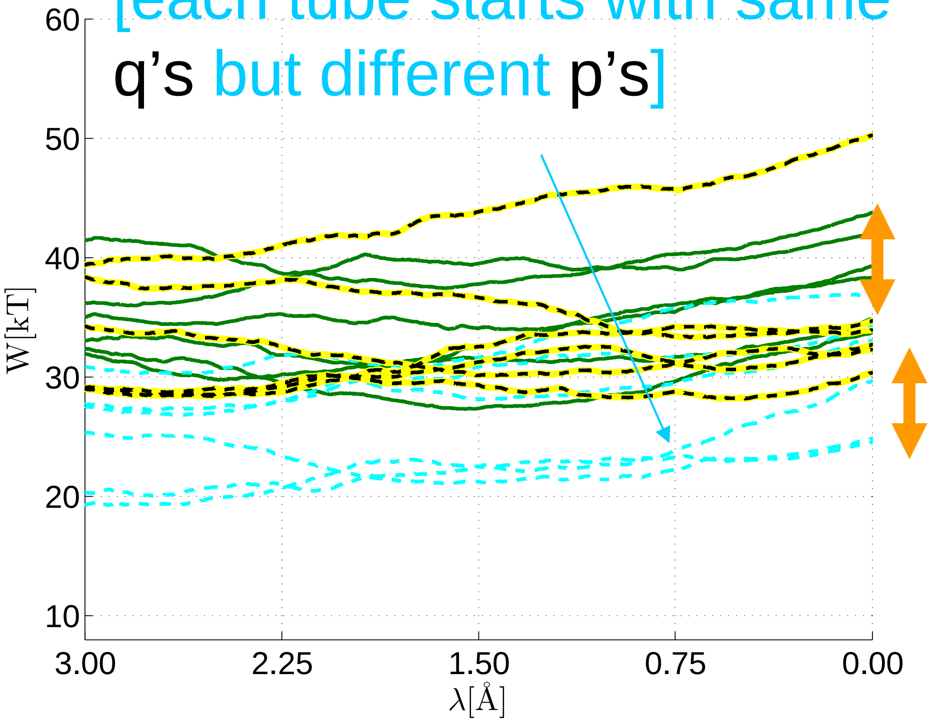
NOTE: Noise Intensity Differs in the Two Sets of Trajectories



PMF and Non-equilibrium Work

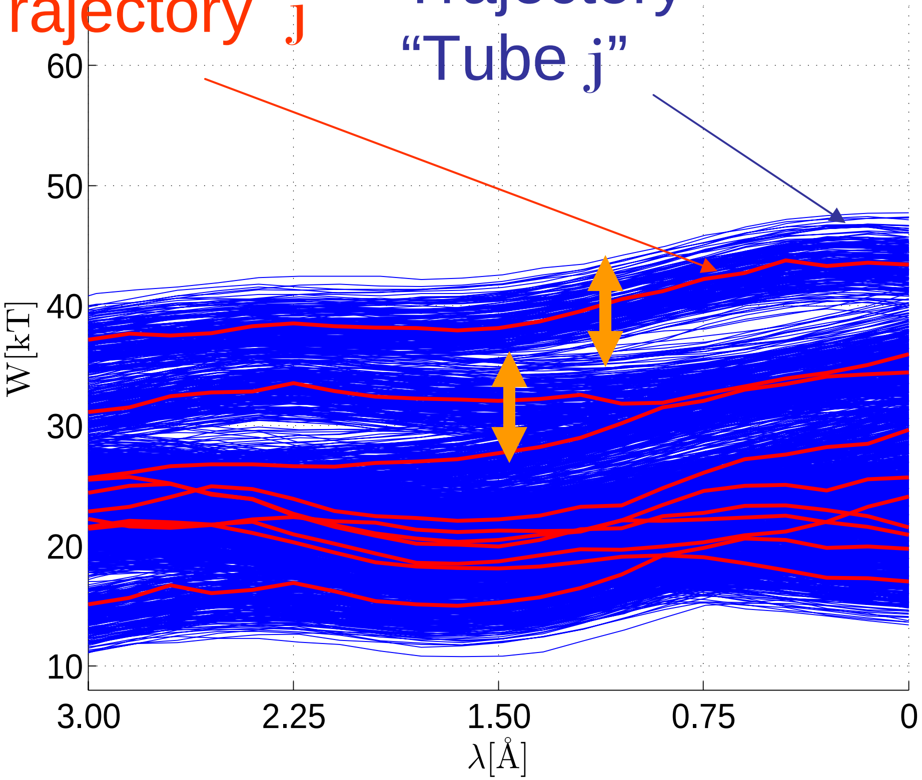
Molecular Dynamics Trajectory "Tube i"

[each tube starts with same q's but different p's]

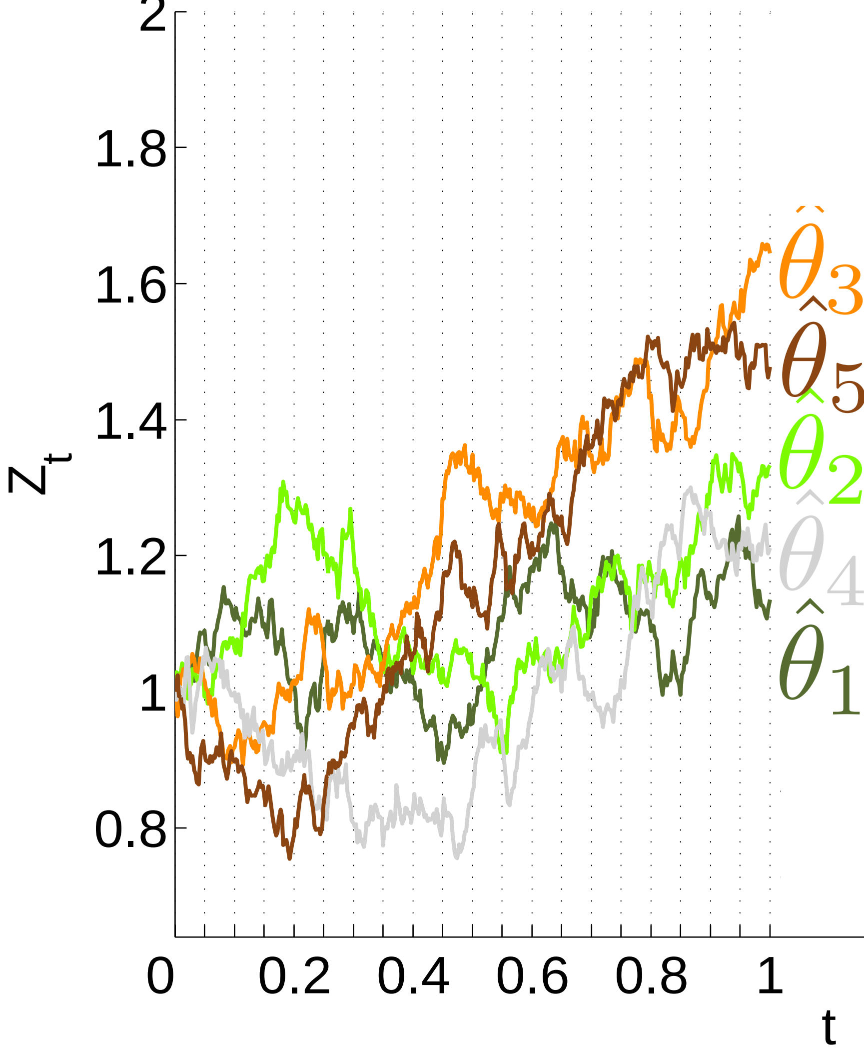


Molecular Dynamics Trajectory "j"

SDE Trajectory "Tube j"



Pathwise View



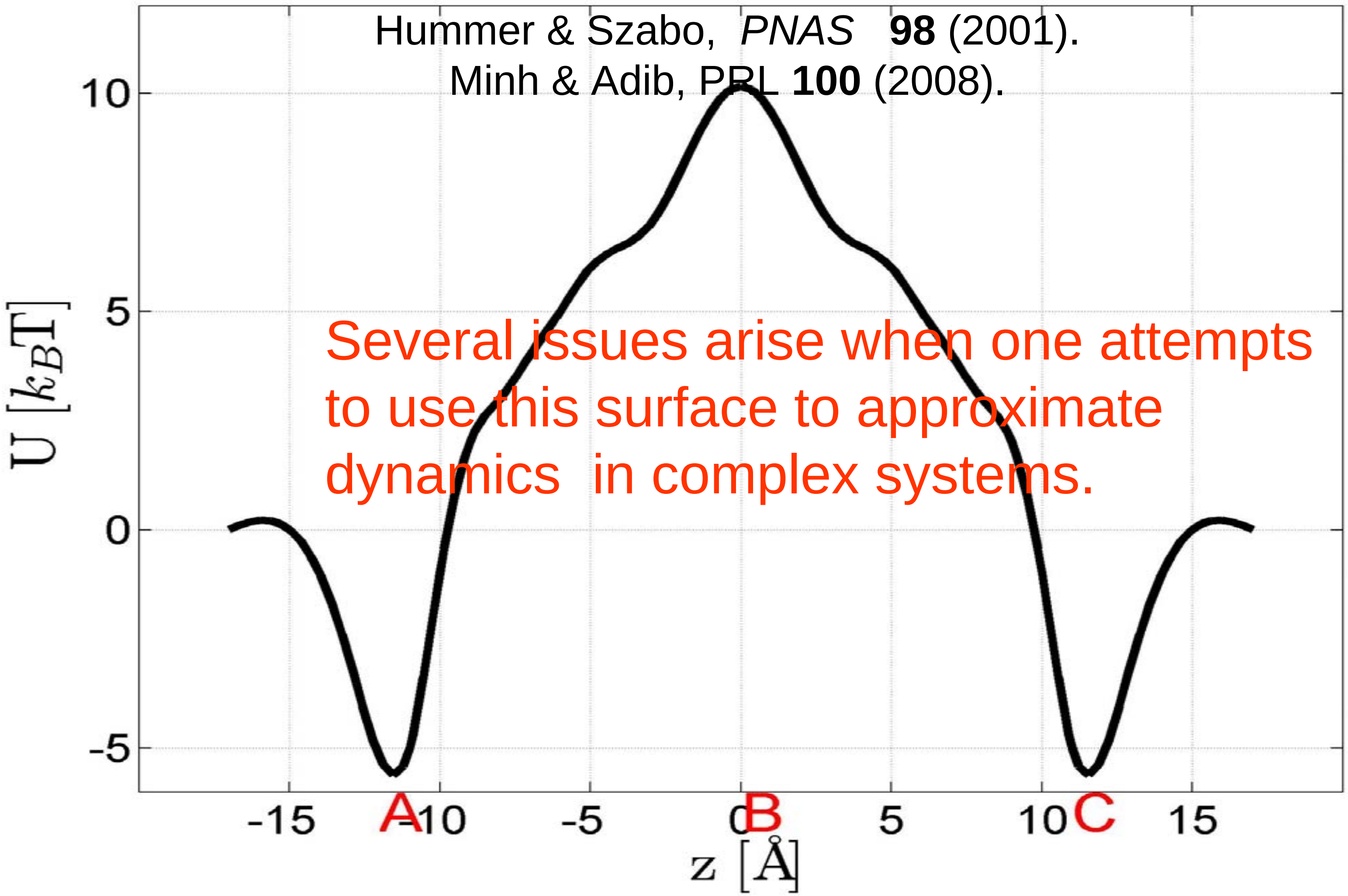
Ensemble Averaging Problematic Due to Nonergodic / Nonstationary Sampling

Collection of Models Another Way of Accounting for "Memory"

$$\cancel{\langle \hat{\theta}_i \rangle}$$

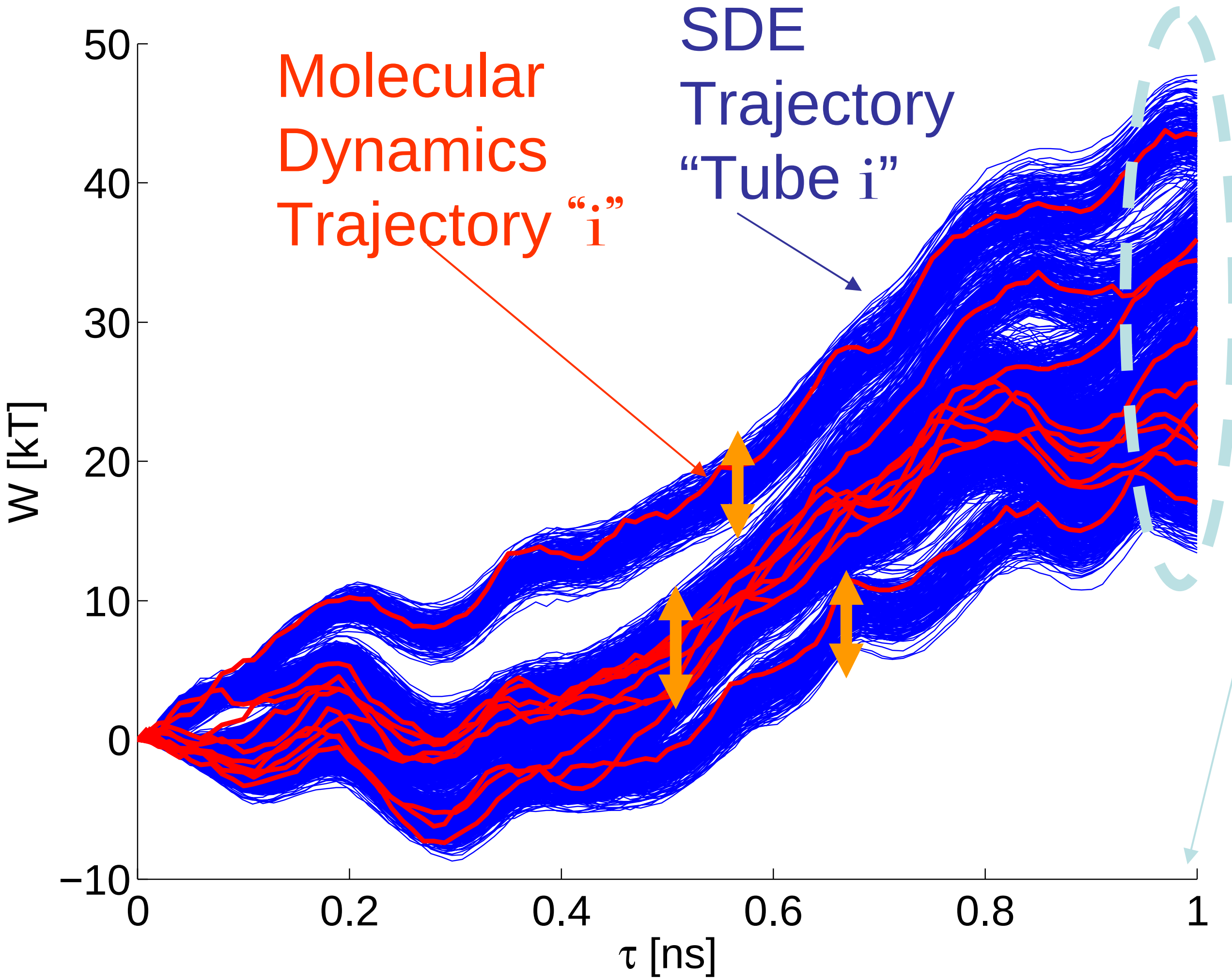
Potential of Mean Force Surface

Hummer & Szabo, *PNAS* **98** (2001).
Minh & Adib, PRL **100** (2008).



Several issues arise when one attempts to use this surface to approximate dynamics in complex systems.

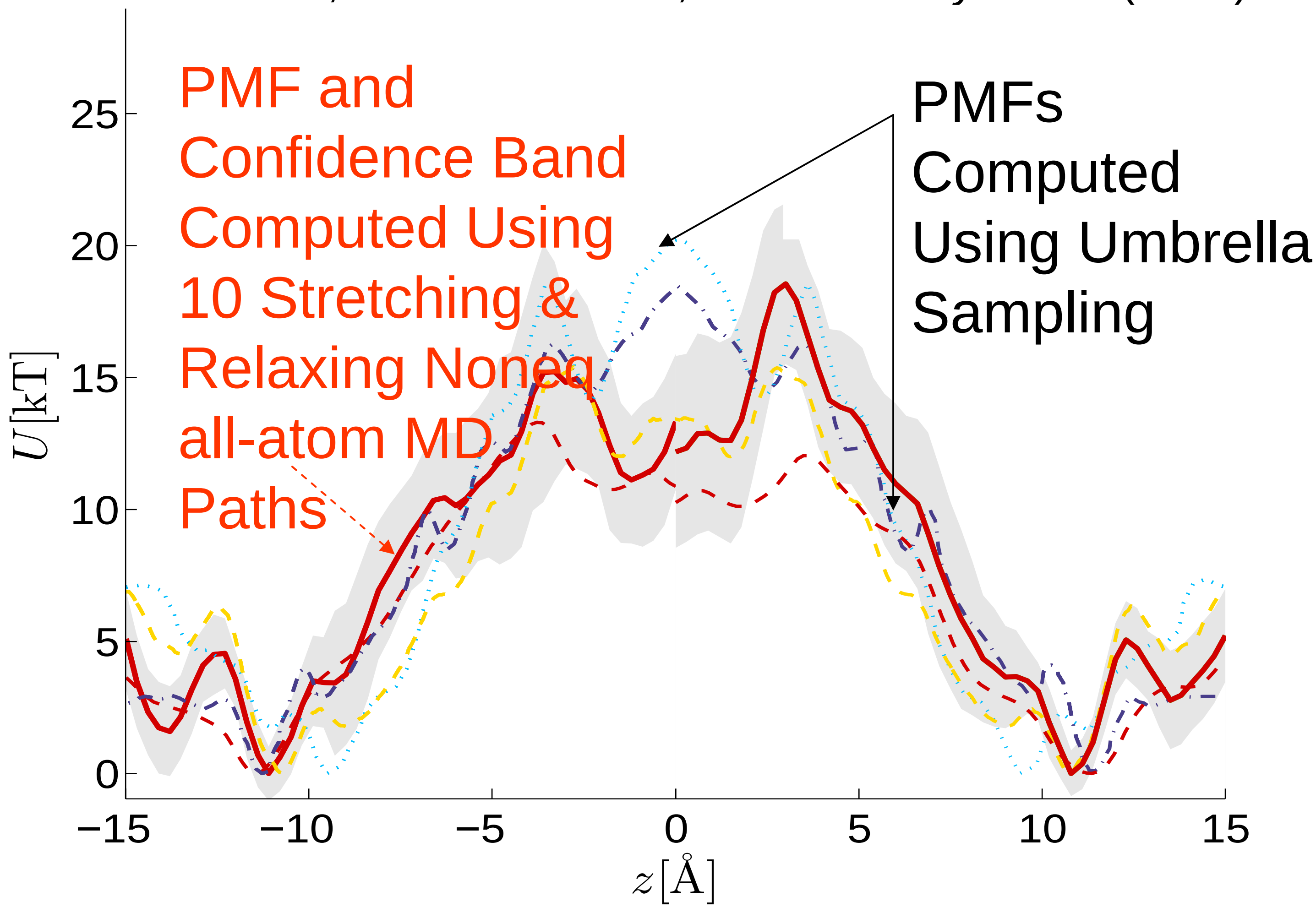
PMF and Non-equilibrium Work



Distribution at Final Time is non-Gaussian and can be viewed as MIXTURE of distributions

PMF of a “Good” Coordinate

Calderon, Janosi & Kosztin, *J. Chem. Phys.* **130** (2009).



Penalized Splines

(See Ruppert, Wand, Carroll's Text: *Semiparametric Regression,* Cambridge University Press, 2003.)

$$y = \{f(x_1), \ldots, f(x_m), \partial f(x_1), \ldots, \partial f(x_m)\} + \epsilon,$$

Observed or Inferred Data

Observation Error

$$y_i = \eta_0 + \eta_1 x_i \ldots \eta_p x_i^p + \sum_{j=1}^{K} \zeta_j B_j(x_i),$$

Spline Basis with K<<m (tall and skinny design matrix)

"Fixed Effects"

"Random Effects"

where $\beta \equiv (\eta, \zeta)$

Penalized Splines

(See Ruppert, Wand, Carroll's Text: *Semiparametric Regression,* Cambridge University Press, 2003.)

$$y = \{f(x_1), \ldots, f(x_m), \partial f(x_1), \ldots, \partial f(x_m)\} + \epsilon,$$

Observed or Inferred Data

$$y_i = \eta_0 + \eta_1 x_i \ldots \eta_p x_i^p + \sum_{j=1}^{K} \zeta_j B_j(x_i),$$

Spline Basis with K<<m (tall and skinny design matrix)

$$\|y - C\beta\|_2^2 + \alpha\|D\beta\|_2^2$$ where $\beta \equiv (\eta, \zeta)$

P-Spline Problem
(Flexible Penalty)

Penalized Splines Using Derivative Information

$$C^{\mathrm{PuDI}} := \begin{pmatrix} 1 & x_1 \dots x_1^p & (\kappa_1 - x_1)_+^p \dots (\kappa_K - x_1)_+^p \\ \vdots & \vdots & \vdots \\ 1 & x_m \dots x_m^p & (\kappa_1 - x_m)_+^p \dots (\kappa_K - x_m)_+^p \\ 0 & 1 \dots p x_1^{p-1} & p(\kappa_1 - x_1)_+^{p-1} \dots p(\kappa_K - x_1)_+^{p-1} \\ \vdots & \vdots & \vdots \\ 0 & 1 \dots p x_m^{p-1} & p(\kappa_1 - x_m)_+^{p-1} \dots p(\kappa_K - x_m)_+^{p-1} \end{pmatrix}.$$

Poorly Conditioned Design Matrix

Pick Smoothness: Solve Many Least Squares Problems

$$\|y - C\beta\|_2^2 + \alpha\|D\beta\|_2^2$$

P-Spline Problem
(Flexible Penalty)

Choose α with Cost Function (GCV). This object has nice "mixed model" interpretation as ratio of observation variance / random effects variance

α Selection Requires Traces and Residuals for Each Candidate

PSQR Penalized Splines using QR

(Calderon, Martinez, Carroll, Sorensen)

Allows Fast and Numerically Stable P-spline Results. Exactly Rank Deficient “C” Can Be Treated

Can Process Many Batches of Curves (Facilitates Solving Many GCV Type Optimization Problems).

Data Mining without User Intervention

(e.g. Closely Spaced and/or Overlapping Knots Are Not Problematic)

Let Regularization Be Handled By Built In Smoothness Penalty (Do Not Introduce Extra Numerical Regularization Steps).

Demmler Reinsch Basis Approach

(Popular Method For Efficiently Computing Smoothing Splines)

Forms Eigenbasis using Cholesky Factor of

$C^T C$

Traces and Residuals for Each Candidate α
Can Easily be Obtained in Vectorized Fashion

Squaring a Matrix is Not a Good Idea (Subtle and Dramatic Consequences)

Avoid ad hoc Regularization

Analogous Existing Methods Do Not Use Good Numerics

e.g. if $C^T C$ Cannot be Cholesky Factored, ad hoc solution:

Find Factor of $C^T C + \eta D$

Instead and Solving that Penalized Regression Problem Originally Posed

Others use SVD truncation in Combination with Regularization (Again Not Problem Posed)

Basic Sketch of Our Approach

(1) Factor C via QR (Done Only Once)

(2) Then Do SVD on Penalized Columns

(3) For Given α Find (Lower Dimensional) QR and Exploit Banded Matrix Structure. Solve Original Penalized Least Squares Problem with **CHEAP QR**

(4) Repeat (3) and Choose α

$$\begin{pmatrix} S \\ \sqrt{\alpha} I \end{pmatrix}$$

Banded "R" Like Matrix

QR and Least Squares

$$\left\| \begin{pmatrix} C \\ \sqrt{\alpha} D \end{pmatrix} \beta - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|_2^2 .$$

Minimizing Vector
Equivalent to Solution of

$$(C^T C + \alpha D^T D)\beta = C^T y$$

$$A^T A \beta = A^T (y^T, 0)^T$$

$$A \equiv \begin{pmatrix} C \\ \sqrt{\alpha} D \end{pmatrix}$$

Efficient QR Treatment?

Exploit P-Spline Problem Structure

Some in Statistics use QR, but Combine Penalized Regression with SVD Truncation

PSQR (Factor Steps)

1. Obtain the QR decomposition of $C = QR$.
2. Partition result above as: $QR = (Q^F, Q^P)\begin{pmatrix} R_{11}^F & R_{12} \\ 0 & R_{22}^P \end{pmatrix}$.
3. Obtain the SVD of $R_{22}^P = USV^T$.
4. Form the following:

$$\tilde{Q} = (Q^F, Q^P U), \tilde{V} = \begin{pmatrix} I & 0 \\ 0 & V^T \end{pmatrix},$$

$$\tilde{R} = \begin{pmatrix} R_{11}^F & R_{12}V \\ 0 & S \end{pmatrix} \equiv \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} \\ 0 & \tilde{R}_{22} \end{pmatrix},$$

$$b = \begin{pmatrix} \tilde{Q}^T y \\ 0 \end{pmatrix} \equiv \begin{bmatrix} \begin{pmatrix} b^F \\ b^P \end{pmatrix} \\ 0 \end{bmatrix}.$$

PSQR (Solution Steps)

5. For each given α (and/or D^P) form: $\widetilde{D}_\alpha = \sqrt{\alpha} D^P V$ and $\widetilde{W}_\alpha = \begin{pmatrix} S \\ \widetilde{D}_\alpha \end{pmatrix}$.
6. Obtain the QR decomposition $\widetilde{W}_\alpha = Q'R'$.
7. Form $c = (R')^{-1}(Q')^T \begin{pmatrix} b^P \\ 0 \end{pmatrix}$.
8. Solve $\hat{\beta}^a_\alpha = \begin{pmatrix} \tilde{R}_{11}^{-1}(b^F - \tilde{R}_{12}c) \\ Vc \end{pmatrix}$.

PSQR (Efficiency)

$$\begin{pmatrix} SV^T \\ D^P \end{pmatrix} = \begin{pmatrix} S \\ D^P V \end{pmatrix} V^T \equiv \begin{pmatrix} \tilde{R}_{22} \\ D^P V \end{pmatrix} V^T,$$

So if $D = diag(0, \ldots, 0, 1, \ldots 1)$
$D^P = diag(1, \ldots 1)$

Then problem reduces to finding x minimizing

$$\left\| \begin{pmatrix} S \\ \sqrt{\alpha} V \end{pmatrix} x - \begin{pmatrix} b^P \\ 0 \end{pmatrix} \right\|_2^2 = \left\| \underbrace{\begin{pmatrix} I, 0 \\ 0, V \end{pmatrix}}_{\text{Orthogonal Matrix}} \underbrace{\begin{pmatrix} S \\ \sqrt{\alpha} I \end{pmatrix}}_{\text{Close to "R"}} x - \begin{pmatrix} b^P \\ 0 \end{pmatrix} \right\|_2^2.$$

PSQR and Givens Rotations

$$D^P = diag(1, \dots 1)$$

$$\left\| \begin{pmatrix} S \\ \sqrt{\alpha} V \end{pmatrix} x - \begin{pmatrix} b^P \\ 0 \end{pmatrix} \right\|_2^2 = \left\| \underbrace{\begin{pmatrix} I, 0 \\ 0, V \end{pmatrix}}_{\textbf{Orthogonal Matrix}} \underbrace{\begin{pmatrix} S \\ \sqrt{\alpha} I \end{pmatrix}}_{\textbf{Close to ``R''}} x - \begin{pmatrix} b^P \\ 0 \end{pmatrix} \right\|_2^2 .$$

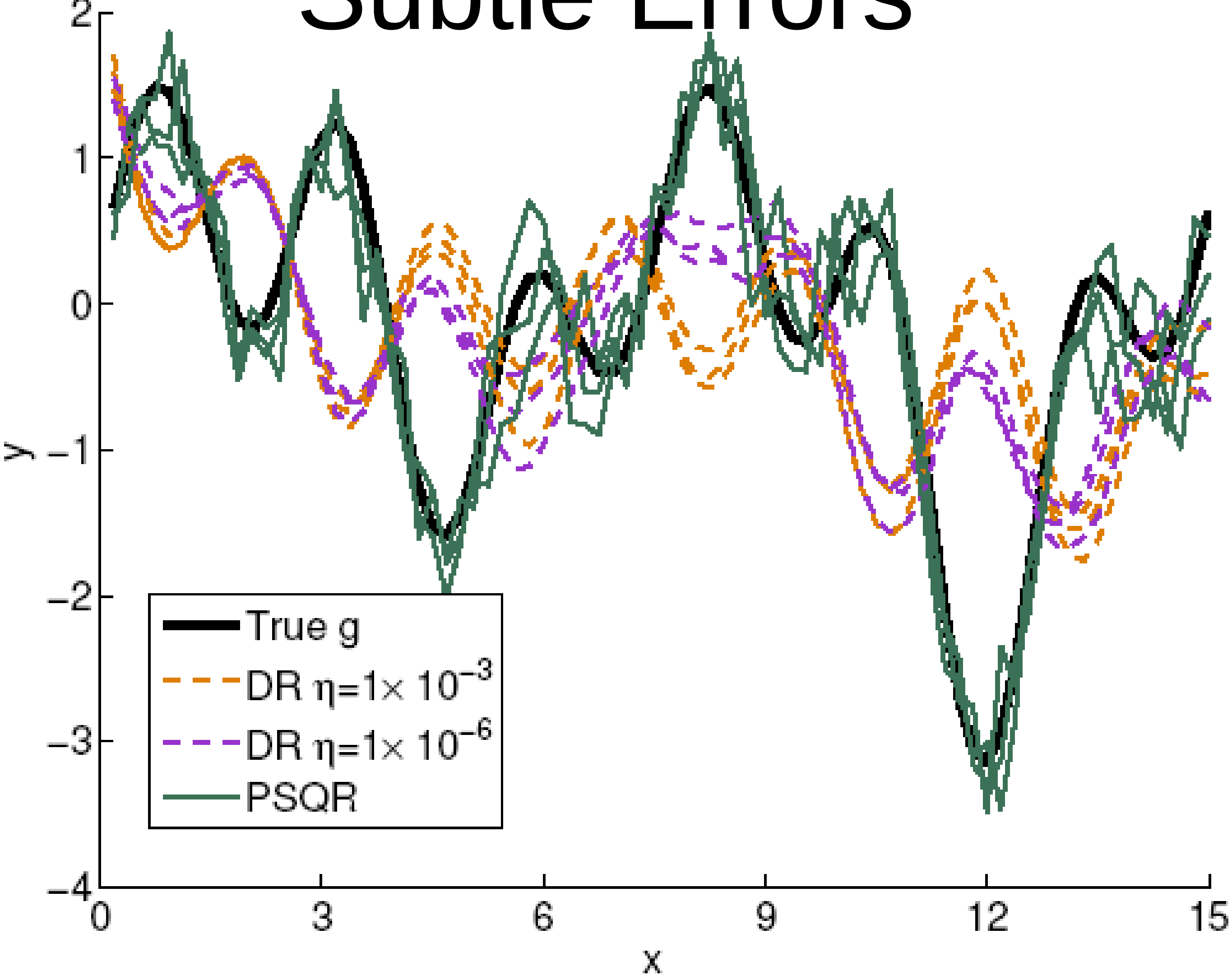
Givens Rotations to Finalize QR

Applying Rotations to RHS Yields:

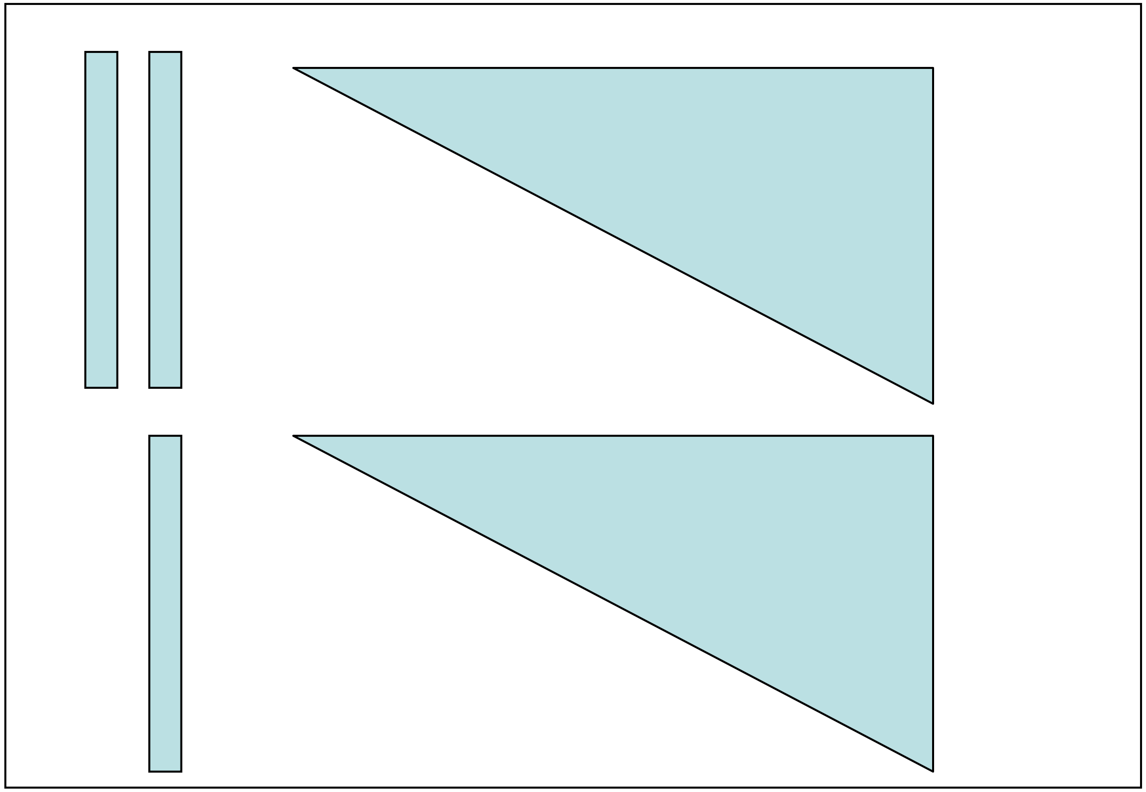
$$(\sqrt{\Lambda})^{-1} S b^P$$

$$\text{Where R} = \sqrt{\Lambda}$$

Subtle Errors



PuDI Design Matrix (One Curve at a Time)



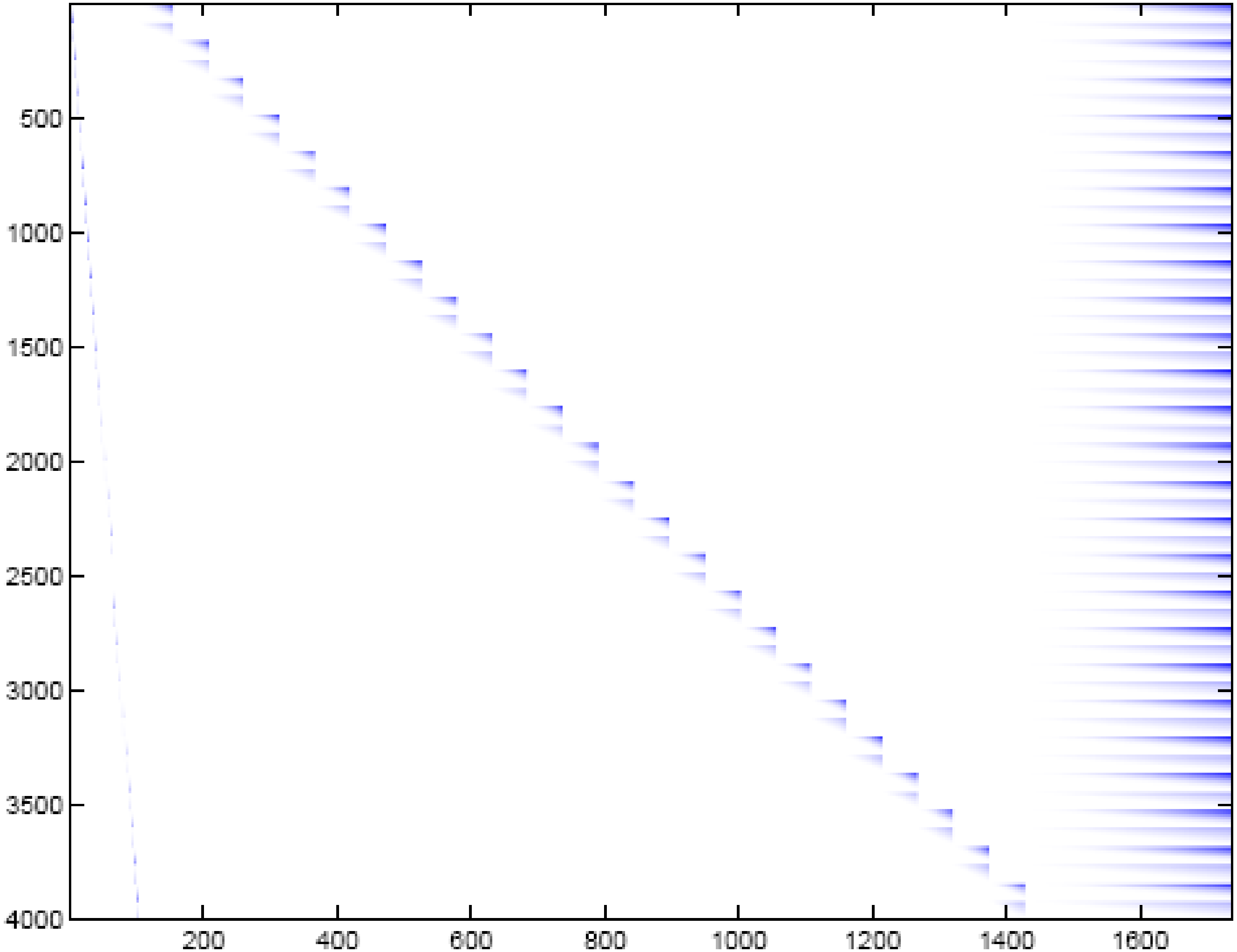
Sparsity Pattern of TPF Basis QR Needed in Step 1 (Before Penalty Added)

TPF: Free and Penalized Blocks

$$C^{\mathrm{PuDI}} := \begin{pmatrix} 1 & x_1 \ldots x_1^p & (\kappa_1 - x_1)_+^p \ldots (\kappa_K - x_1)_+^p \\ \vdots & \vdots & \vdots \\ 1 & x_m \ldots x_m^p & (\kappa_1 - x_m)_+^p \ldots (\kappa_K - x_m)_+^p \\ 0 & 1 \ldots p x_1^{p-1} & p(\kappa_1 - x_1)_+^{p-1} \ldots p(\kappa_K - x_1)_+^{p-1} \\ \vdots & \vdots & \vdots \\ 0 & 1 \ldots p x_m^{p-1} & p(\kappa_1 - x_m)_+^{p-1} \ldots p(\kappa_K - x_m)_+^{p-1} \end{pmatrix}.$$

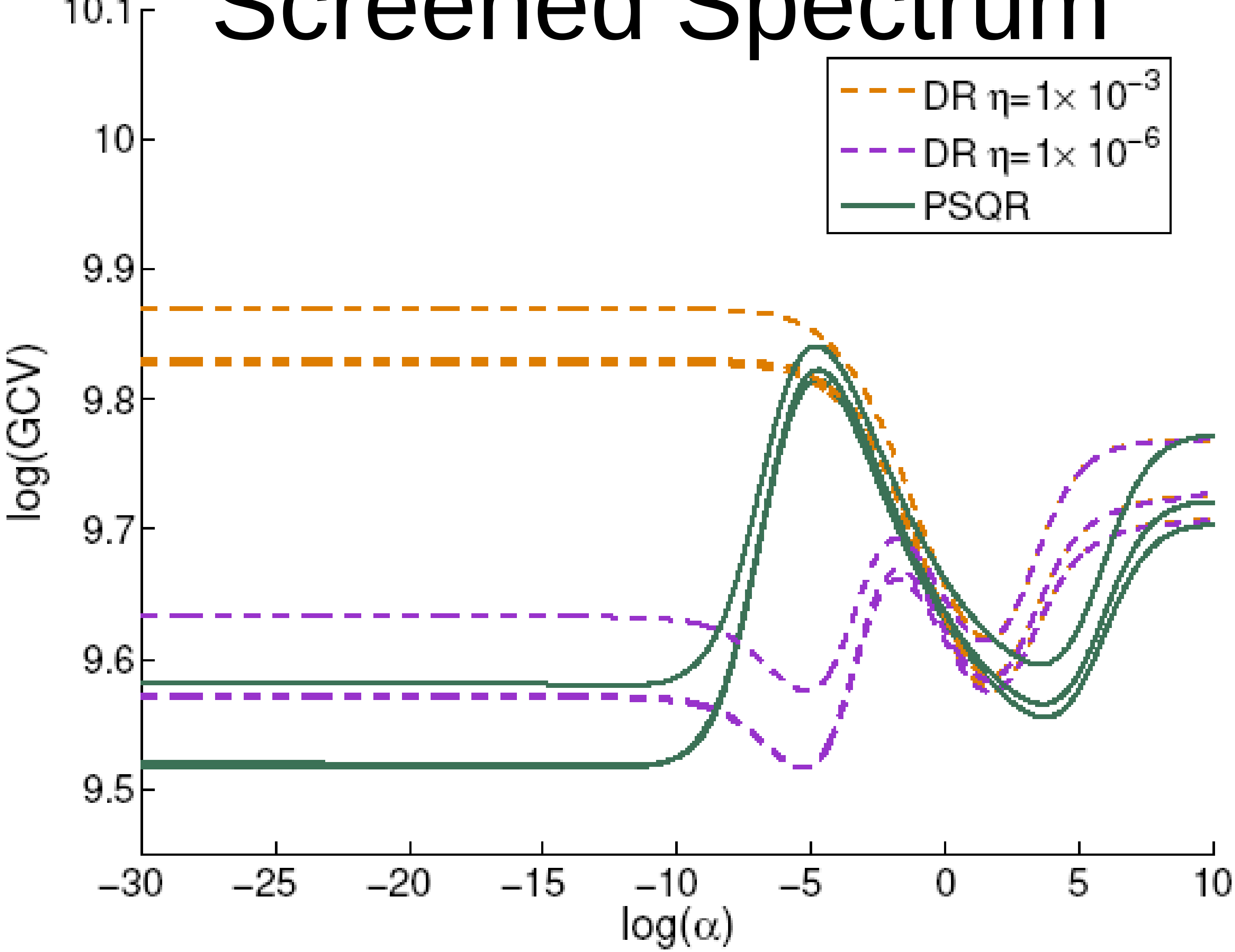
Last K Columns

PuDI Design Matrix (Batches of Curves)



Sparsity Pattern of TPF Basis QR Needed in Step 1 (Before Penalty Added)

Screened Spectrum

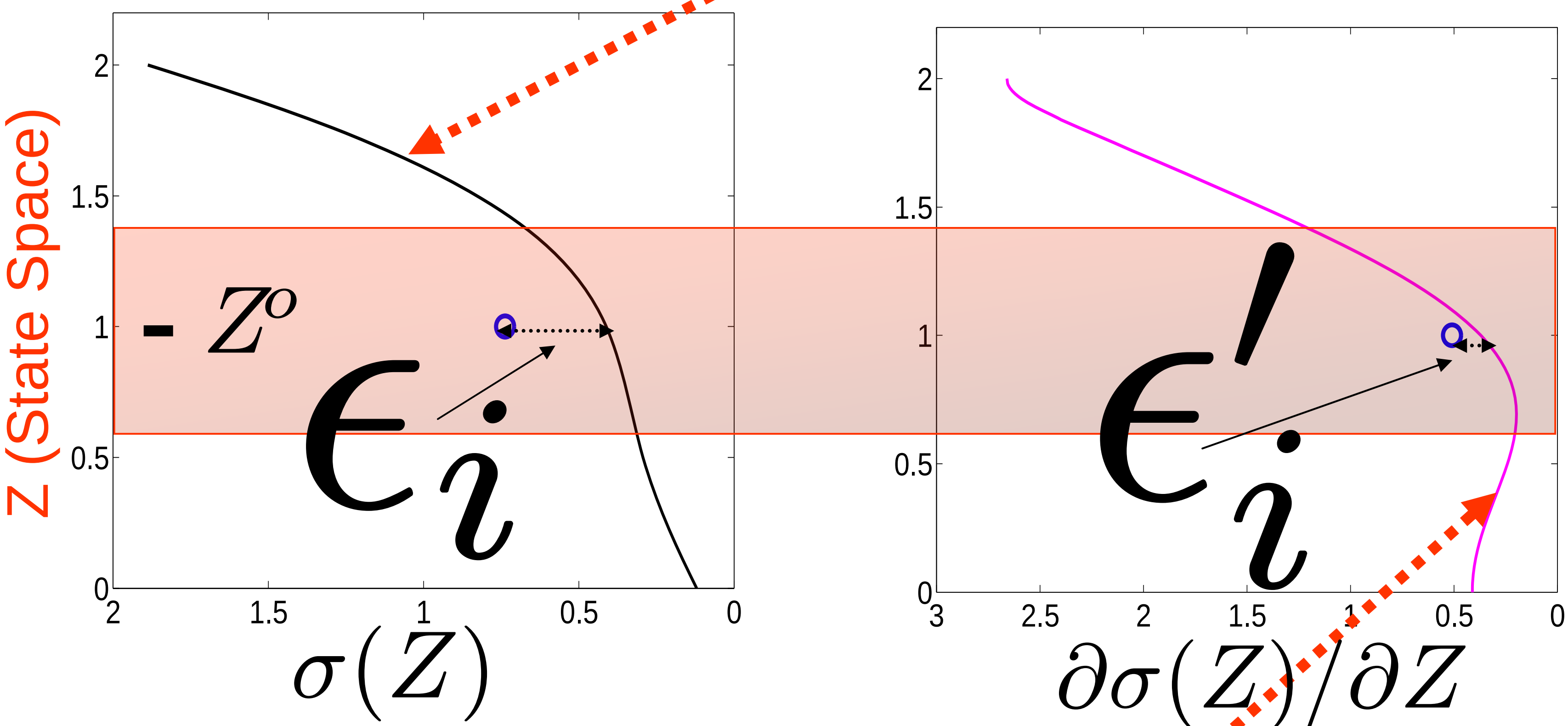


Simple Basis Readily Allows For Useful Extensions

(e.g. Generalized Least Squares)

Function of Interest (the “truth”)

Noisy Point Estimates
(finite discrete time series sample uncertainty)



Spatial Derivative of Function of Interest

Point Estimates Noise Distribution Depends on Window and Function Estimated (Quantifying and Balancing These Can Be Important); Especially for Resolving "Spiky" Features.