International Journal of Mathematics, Statistics and Operations Research
Vol. 2; Number 2; 2022, pp. 79-99
https://doi.org/10.47509/IJMSOR.2022.v02i02.01

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

Optimal Estimation of Brownian Penalized Regression Coefficients

Paramahansa Pramanik [1] **and Alan M. Polansky** [2]

[1] *Department of Mathematics and Statistics*
University of South Alabama
Mobile, AL 36688 USA.

[2] *Department of Statistics and Actuarial Science*
Northern Illinois University
DeKalb, IL 60115 USA.

Abstract: In this paper we introduce a new methodology to determine an optimal coefficient of penalized functional regression. We assume the dependent, independent variables and the regression coefficients are functions of time and error dynamics follow a stochastic differential equation. First we construct our objective function as a time dependent residual sum of square and then minimize it with respect to regression coefficients subject to different error dynamics such as LASSO, group LASSO, fused LASSO and cubic smoothing spline. Then we use Feynman-type path integral approach to determine a Schrödinger-type equation which have the entire information of the system. Using first order conditions with respect to these coefficients give us a closed form solution of them.

MSC 2010 subject classifications: Primary 60H05; Secondary 81Q30.
Keywords and phrases: Penalized regression coefficients, Feynman-type path integrals, stochastic differential equations.

1. Introduction

Functional regression has been popular in recent times (Ramsay, 2004; Ramsay and Silverman, 2007). Different penalizations such as least absolute shrinkage and selection operator (LASSO), ridge regression, standard L^p-norm, elastic net regression, Group LASSO, fused LASSO, bridge regression and different types of splines have been used in statistical literature for variable selection. Following Huang and Su (2021) we know, penalized regression has been popularized after publication of Eilers and Marx (1996) and Ruppert, Wand and Carroll (2003). A mean squared error of penalized spline estimators under a white noise model was obtained in Hall and Opsomer (2005). Furthermore, different approximation of penalized spline estimators have been done in Li and Ruppert (2008), Wang, Shen and Ruppert (2011), Schwarz and Krivobokova (2016) and Lai and Wang (2013). These works have been used in closed-form expressions of penalized spline estimators which are only available in the regression setting where all the variables are time independent. When such expressions are not available in other estimation contexts, such as estimation of density functions or conditional quantile functions, Huang and Su (2021) or furthermore, when the penalization function is itself a stochastic differential equation, existing asymptotic approaches extended. Then we need a path integral approach to determine regression coefficients in Euclidean field (Pramanik, 2020; Pramanik and Polansky, 2020a,b; Pramanik, 2021a,b) and for generalized tensor field (Pramanik and Polansky, 2019).

In this paper we provide a dynamic framework of a time dependent residual sum of square and minimize it with respect to regression coefficients where coefficient dynamics follow a stochastic differential equation. We construct a quantum Lagrangian for equal in length small time interval with respect to a positive penalization parameter and use a Feynman-type path integral approach to determine a Schrödinger type equation (Pramanik, 2016; Hua, Polansky and Pramanik, 2019; Pramanik, 2020, 2021a; Polansky and Pramanik, 2021) and optimal values of the regression coefficients are the first order condition of it (Baaquie, 2007; Feynman, 1949;

Article History
Received : 25 September 2022; Revised : 08 November 2022; Accepted : 08 November 2022; Published : 29 December 2022

To cite this paper
Paramahansa Pramanik & Alan M. Polansky (2022). Optimal Estimation of Brownian Penalized Regression Coefficients. *International Journal of Mathematics, Statistics and Operations Research.* 2(2), 79-99.

Pramanik, 2021c) and Fujiwara (2017); Pramanik and Polansky (2021). As at the beginning of a new time interval we do not have any prior knowledge about the future, a conditional expectation until that initial time point of the residual sum of squares is used as our objective function. In examples we show closed form of the regression coefficients under different penalizations. Traditional literature of this type regression does not consider diffusion part of the process. Hence, we cannot see the bigger picture of it and more generalization towards Brownian motion is needed.

Before constructing the quantum Lagrangian for small time intervals and path integral of the system we showed those two integrals exist under certain assumptions, which will be discussed in the next section. Main motivation of using Feynman path integral approach is it considers all possible paths between two time points and eliminates the extremes by Lebesgue-Riemann lemma to determine the minimized action locally. Furthermore, this approach gives solution for more generalized system of equations where Pontryagin's optimal principle fails (Baaquie, 2007; Bellman, 1966) and Yeung and Petrosjan (2006).

2. Preliminaries

Consider a sample of N time dependent cases each of which consists of J covariates such that for an observation i we have the following regression model

$$Y_i(s) = \sum_{j'=1}^{J} \beta_{j'}(s) X_{ij'}(s) + U_i(s),$$

where $\beta_{j'}(s) \in \boldsymbol{\beta}(s) \in \mathbb{R}^J$ for all $j' = 1, ..., J$, $Y_i(s) \in \mathbb{R}$ is i^{th} outcome and $X_{ij'}(s) \in \mathbb{R}$ is i^{th} independent variable corresponding to deterministic $\beta_{j'}(s)$ coefficient, with $i = 1, ..., N$ and time $s \in [0, T]$ and, the error term $U_i \in \mathbf{U} \in \mathbb{R}^N$ is assumed to be a stochastic process expressed by the stochastic differential Equation (1) below.

Therefore, to obtain an optimal regression coefficient the objective is to minimize time dependent residual sum of square (RSS)

$$\overline{\mathbf{X}}_O(s, \boldsymbol{\beta}, \mathbf{X}) = \sum_{i=1}^{N} \left[Y_i(s) - \sum_{j'=1}^{J} \beta_{j'}(s) X_{ij'}(s) \right]^2,$$

with respect to $\beta_{j'}(s) \in \boldsymbol{\beta}(s) \in \mathbb{R}^J$ Furthermore, we assume the N-dimensional error vector $\mathbf{U}(\mathbf{s})$ follows a stochastic differential equation,

$$d\mathbf{U}(s) = \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]ds + \boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]d\mathbf{B}(s), \tag{1}$$

where $\boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]$ is a $N \times 1$-dimensional drift vector, $\boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]$ a $N \times p$-dimensional diffusion matrix and $\mathbf{B}(s)$ is a p-dimensional Brownian motion. The mappings of $\boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]$ and $\boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]$ are jointly measurable and continuous. For $s \in [0, T]$ the mapping $\boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] : C^0([0, T], \mathbb{R}^J, \mathbb{R}^{N \times J}) \to L(\mathbb{R}^n, \mathbb{R}^J, \mathbb{R}^{N \times J})$ and $\boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] : C^0([0, T], \mathbb{R}^J, \mathbb{R}^{N \times J}) \to L(\mathbb{R}^n, \mathbb{R}^J, \mathbb{R}^{N \times J})$ are measurable with respect to the σ-algebra generated by the cylindrical sets with bases over the the time interval $[0, T]$ in continuous function vanishing at the infinity $C^0([0, T], \mathbb{R}^J, \mathbb{R}^{N \times J})$, and the Borel σ-algebras in $\mathbb{R}^J$, $\mathbb{R}^{N \times J}$ and a linear functional $L(\mathbb{R}^n, \mathbb{R}^J, \mathbb{R}^{N \times J})$ on a filtration $\mathcal{F}_s$ starting at time s, where time interval $[0, T]$ has been divided into n small equal-lengthed subintervals. If above conditions hold, then for initial condition $\mathbf{X}_0 \in \mathbb{R}^{(N \times J) \times 1}$ Krylov's theorem tells that, there exists a weak solution of coefficient dynamics represented by the Equation (1) Krylov (2008). The drift coefficient $\boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]$ of the coefficient dynamics have different forms like for LASSO with m covariates it is $\sum_{j'=1}^{m} |\beta_{j'}(s)|$, ridge regression $\sum_{j'=1}^{m}\beta_{j'}^2(s)$, standard L^p- norm $[\sum_{j'=1}^{m}|\beta_{j'}(s)|^p]^{(1/p)}$, elastic net regression $(1-\alpha)||\boldsymbol{\beta}(s)||_1+\alpha||\boldsymbol{\beta}(s)||_2^2$ with $\alpha\in[0,1]$, group LASSO $\sum_{j'=1}^{m}\beta_{j'}^T(s)K_{j'}(s)\beta_{j'}(s)$ with $K_{j'}$ being a positive definite matrix, fused LASSO $\alpha\sum_{j'=1}^{m}|\beta_{j'}(s)|+(1-\alpha)\sum_{j'=1}^{m}|\beta_{j'}(s)-\beta_{j'-1}(s)|$ and bridge regression $(\sum_{j'=1}^{m}\sqrt{|\beta_{j'}(s)|})^2$ which we will discuss in examples. Furthermore, as we are concentrating in dynamic optimization, our objective is to

$$\min_{\{\beta'_j\in\boldsymbol{\beta}\}}\overline{\mathbf{X}}_O(s,\boldsymbol{\beta},\mathbf{X})=\min_{\{\beta'_j\in\boldsymbol{\beta}\}}\mathbb{E}\int_0^T\sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]^2 ds, \tag{2}$$

subject to the Equation (1). To solve for the optimal coefficients we use Feynman-type path integral approach Feynman (1949) where we define a quantum Lagragian action function for small time interval $[s,\tau]\subseteq[0,T]$ as

$$\mathcal{L}_{s,\tau}(\mathbf{X})=\mathbb{E}_s\int_s^\tau\Bigg\{\sum_{i=1}^{N}\left[Y_i(\nu)-\sum_{j'=1}^{J}\beta_{j'}(\nu)X_{ij'}(\nu)\right]^2 d\nu \\ +\lambda[\boldsymbol{\mu}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]d\nu+\boldsymbol{\sigma}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]d\mathbf{B}(\nu)-\Delta\mathbf{U}(\nu)]\Bigg\}, \tag{3}$$

where $\lambda>0$ is the time independent penalization parameter. We will show the above integral in Equation (3) measurable and then Feynman path integral of it is also measurable in $\mathbb{R}^{N\times J}$ Feynman (1949). Later part of this paper in Proposition 2 we will discuss about the closed form solutions of these coefficients under smoothing spline environment.

3. Definitions and Assumptions

Definition 1. *Suppose a space $\mathcal{X}$ is Hausdorff. If for every point $x\in\mathcal{X}$ and every closed set $\mathcal{Z}\subseteq\mathcal{X}$ not containing x, there exists a continuous function $g_c:\mathcal{X}\to[0,1]$ such that, $g_c(x)=1$ and $g_c(z)=0$ for all $z\in\mathcal{Z}$ then, $\mathcal{X}$ is completely regular Bogachev (2007).*

Definition 2. *For a family $\mathcal{M}$ of Radon measures on a topological space $\mathcal{X}$ if for every $\varepsilon>0$, there exists a compact set κ_ε such that $|\rho|(\mathcal{X}\setminus\kappa_\varepsilon)<\varepsilon$ for all $\rho\in\mathcal{M}$ then $\mathcal{X}$ is called uniformly tight Bogachev (2007).*

Furthermore, from Definition 2 and Prohorov Theorem we know, if $\mathcal{M}$ is a family of Borel measures on $\mathcal{X}$ then every sequence $\{\rho_n\}_{n\geq1}\subset\mathcal{M}$ contains a weakly convergent subsequence or $\mathcal{M}$ is uniformly tight and bounded (Bogachev, 2007; Prokhorov, 1956). In order to understand projective system of spaces let us assume $\mathcal{T}$ be a directed set and let $\{\mathbf{X}_n\}_{n\in\mathcal{T}}$ with γ be a continuous mapping such that for two indices $n\geq m$ the condition $\gamma_{mn}:\mathbf{X}_n\to\mathbf{X}_m$ and for $\eta\geq n\geq m$, $\gamma_{mn}\circ\gamma_{n\eta}=\gamma_{m\eta}$ hold. Furthermore, suppose $\mathbf{X}$ be a space such that mapping $\gamma_m:\mathbf{X}\to\mathbf{X}_m$ is consistent with γ_{nm} by the mapping $\gamma_m=\gamma_{mn}\circ\gamma_n$ for all $m\leq n$. Then $\mathbf{X}_m$ is the inverse limit space. As $\mathbf{X}=\mathbb{R}^\infty$ is an example of this space, the dimension of our independent variables $\mathbf{X}_{nJ}=\mathbb{R}^{n\times J}$ consists of all sequences of the form $(X_1,...,X_{nJ},0,...,0)$, and γ_{nJk} and γ_{nJ} are natural projections. Now consider spaces $\mathbf{X}_n$ are equipped with Borel σ-algebra $\mathcal{B}_n$ and measures ρ_n on $\mathcal{B}_n$ such that γ_{mn} are measurable. Then for $m\leq n$

$$\gamma_{mn}(\rho_n):=\rho_n\circ\gamma_{mn}^{-1}=\rho_m,$$

is a necessary condition. Furthermore, for ρ is a Radon measure on $\mathbf{X}$, $\rho\circ\gamma_n^{-1}=\rho_n$, $\forall n$ exists iff for any $\varepsilon>0$, $\exists\kappa_\varepsilon\subset\mathbf{X}$ with $\rho_n(\gamma_n(\kappa_\varepsilon))\geq1-\varepsilon$, $\forall n$ Bogachev (2007). We use this result to prove Lemma 1.

Assumption 1. *For time interval $[s, s+\varepsilon] \subset [0, T]$, where $\varepsilon \downarrow 0$ the filtration space starting at time s denoted by $\mathcal{F}_s$, is a vector lattice of on the non-empty set Ω such that for point-transition $\Psi_s(\mathbf{X})$ and quantum Lagrangian $\mathcal{L}_{s,s+\varepsilon}$,*

$$\Psi_{s,s+\varepsilon}(\mathbf{X}) = \frac{1}{N_s^{\tilde{f}}} \int_{\mathbb{R}^{N \times J}} \tilde{f} d\mathbf{X},$$

where $\tilde{f} = \exp\left[-\varepsilon \mathcal{L}_{s,s+\varepsilon}(\mathbf{X})\right] \Psi_s(\mathbf{X})$ and $N_s^{\tilde{f}} > 0$ is a normalizing constant of $\tilde{f} \in \mathcal{F}_s$. For another function $\tilde{g} \in \mathcal{F}_s$ with normalizing constant $N_s^{\tilde{g}} > 0$ define

$$\tilde{\Psi}_{s,s+\varepsilon}(\mathbf{X}) = \frac{1}{N_s^{\tilde{g}}} \int_{\mathbb{R}^{N \times J}} \tilde{g} d\mathbf{X},$$

such that $\max(\tilde{f}, \tilde{g}) \in \mathcal{F}_s$, $\min(\tilde{f}, \tilde{g}) = -\max(\tilde{f}, \tilde{g})$ *and* $|\tilde{f}| \in \mathcal{F}_s$.

Assumption 2. *The set of all bounded functions $\mathcal{F}_s^+$ of $\tilde{f}$ such that for a non-negative increasing sequence $\tilde{f}_k \in \mathcal{F}_s$ the condition $\tilde{f} = \lim_{k\to\infty} \tilde{f}_k$ holds. As the sequence $\{\tilde{f}_k\}$ is uniformly bounded, we assume the sequence $\{\Psi_{s,s+\varepsilon}^k\}$ is increasing and bounded where,*

$$\Psi_{s,s+\varepsilon}^k(\mathbf{X}) = \frac{1}{N_s^{\tilde{f}_k}} \int_{\mathbb{R}^{N \times J}} \tilde{f}_k d\mathbf{X}.$$

Assume $\Psi_{s,s+\varepsilon}(\mathbf{X}) = \lim_{k\to\infty} \Psi_{s,s+\varepsilon}^k(\mathbf{X})$. Then For all $\tilde{f}, \tilde{g} \in \mathcal{F}_s^+$ and $\tilde{f} \leq \tilde{g}$ there exists a measure $(N_s^{\tilde{f}})^{-1} d\mathbf{X}$ such that following conditions hold,

1. $\Psi_{s,s+\varepsilon}(\mathbf{X}) \leq \tilde{\Psi}_{s,s+\varepsilon}(\mathbf{X})$;
2. $\Psi_{s,s+\varepsilon}^*(\mathbf{X}) = \Psi_{s,s+\varepsilon}(\mathbf{X}) + \tilde{\Psi}_{s,s+\varepsilon}(\mathbf{X})$, *where*

$$\Psi_{s,s+\varepsilon}^*(\mathbf{X}) = \frac{1}{N_s^{(\tilde{f}+\tilde{g})}} \int_{\mathbb{R}^{N \times J}} (\tilde{f} + \tilde{g}) d\mathbf{X}.$$

3. *For a constant $c \in [0, \infty)$, $\Psi_{s,s+\varepsilon}^c(\mathbf{X}) = c\Psi_{s,s+\varepsilon}(\mathbf{X})$ where,*

$$\Psi_{s,s+\varepsilon}^c(\mathbf{X}) = \frac{1}{N_s^{(c\tilde{f})}} \int_{\mathbb{R}^{N \times J}} (c\tilde{f}) d\mathbf{X}.$$

4. *For all $\min(\tilde{f}, \tilde{g}) \in \mathcal{F}_s^+$ and $\max(\tilde{f}, \tilde{g}) \in \mathcal{F}_s^+$ we have $\Psi_{s,s+\varepsilon}(\mathbf{X}) + \tilde{\Psi}_{s,s+\varepsilon} = \Psi_{s,s+\varepsilon}^{\min}(\mathbf{X}) + \Psi_{s,s+\varepsilon}^{\max}(\mathbf{X})$ where*

$$\Psi_{s,s+\varepsilon}^{\min}(\mathbf{X}) = \frac{1}{N_s^{\min(\tilde{f},\tilde{g})}} \int_{\mathbb{R}^{N \times J}} \min(\tilde{f}, \tilde{g}) d\mathbf{X}$$

and

$$\Psi_{s,s+\varepsilon}^{\max}(\mathbf{X}) = \frac{1}{N_s^{\max(\tilde{f},\tilde{g})}} \int_{\mathbb{R}^{N \times J}} \max(\tilde{f}, \tilde{g}) d\mathbf{X}.$$

5. $\lim_{k\to\infty} \tilde{f}_k \in \mathcal{F}_s^+$ *for every uniformly bounded sequence of $\tilde{f}_k \in \mathcal{F}_s^+$, and one has $\Psi_{s,s+\varepsilon}^{\lim}(\mathbf{X}) = \lim_{k\to\infty} \Psi_{s,s+\varepsilon}^k(\mathbf{X})$ where*

$$\Psi_{s,s+\varepsilon}^{\lim}(\mathbf{X}) = \frac{1}{N_s^{\lim_{k\to\infty} \tilde{f}_k}} \int_{\mathbb{R}^{N \times J}} \lim_{k\to\infty} \tilde{f}_k d\mathbf{X}.$$

Assumption 3. *For $T > 0$, let $\boldsymbol{\mu}(s, \boldsymbol{\beta}, \mathbf{X}) : C^0([0, T], \mathbb{R}^J, \mathbb{R}^{N \times J}) \to L(\mathbb{R}^n, \mathbb{R}^J, \mathbb{R}^{N \times J})$ and $\boldsymbol{\sigma}(s, \boldsymbol{\beta}, \mathbf{X}) : C^0([0, T], \mathbb{R}^J, \mathbb{R}^{N \times J}) \to L(\mathbb{R}^n, \mathbb{R}^J, \mathbb{R}^{N \times J})$ be some measurable function and, for some positive constant K_1 and, $\mathbf{X} \in \mathbb{R}^{N \times J}$ we have linear growth of $\boldsymbol{\beta}$ as*

$$|\boldsymbol{\mu}(s, \boldsymbol{\beta}, \mathbf{X})| + |\boldsymbol{\sigma}(s, \boldsymbol{\beta}, \mathbf{X})| \leq K_1(1 + |\mathbf{X}|),$$

such that, there exists another positive, finite, constant K_2 and for a different vector $\widetilde{\mathbf{X}}_{(N\times J)\times 1}$ such that the Lipschitz condition,

$$|\boldsymbol{\mu}(s,\boldsymbol{\beta},\mathbf{X})-\boldsymbol{\mu}(s,\boldsymbol{\beta},\widetilde{\mathbf{X}})|+|\boldsymbol{\sigma}(s,\boldsymbol{\beta},\mathbf{X})-\boldsymbol{\sigma}(s,\boldsymbol{\beta},\widetilde{\mathbf{X}})|\leq K_2\ |\mathbf{X}-\widetilde{\mathbf{X}}|,$$

$\widetilde{\mathbf{X}}\in\mathbb{R}^{N\times J}$ *is satisfied and*

$$|\boldsymbol{\mu}(s,\boldsymbol{\beta},\mathbf{X})|^2+\|\boldsymbol{\sigma}(s,\boldsymbol{\beta},\mathbf{X})\|^2\leq K_2^2(1+|\widetilde{\mathbf{X}}|^2),$$

where $\|\boldsymbol{\sigma}(s,\boldsymbol{\beta},\mathbf{X})\|^2=\sum_{i=1}^N\sum_{j=1}^N|\sigma^{ij}(s,\boldsymbol{\beta},\mathbf{X})|^2$.

Assumption 4. *There exists a probability space $(\Omega,\mathcal{F}_s^{\mathbf{X}},\mathcal{P})$ with sample space Ω, filtration at time s of independent variable $\mathbf{X}$ as $\{\mathcal{F}_s^{\mathbf{X}}\}\subset\mathcal{F}_s$, a probability measure $\mathcal{P}$ and a p-dimensional $\{\mathcal{F}_s\}$ Brownian motion $\mathbf{B}$ where the measure of the regression coefficient $\boldsymbol{\beta}$ is an $\{\mathcal{F}_s^{\mathbf{X}}\}$ adapted process such that Assumption 3 holds.*

4. Main Results

The objective function is,

$$\min_{\{\beta_j'\in\boldsymbol{\beta}\}}\overline{\mathbf{X}}_O(s,\boldsymbol{\beta},\mathbf{X})=\min_{\{\beta_j'\in\boldsymbol{\beta}\}}\mathbb{E}\int_0^T\sum_{i=1}^N\left[Y_i(s)-\sum_{j'=1}^J\beta_{j'}(s)X_{ij'}(s)\right]^2ds, \tag{4}$$

In Equation (4), β_j is the coefficient of independent variable X_{ij} for all $i=1,...,N$ and $j'=1,...,J$.

Lemma 1. *Suppose time interval $[0,T]$ and $\mathbf{R}^{N\times J}$ are completely regular space such that the space $\mathcal{T}=[0,T]\times\mathbb{R}^{N\times J}$ is also completely regular and all the compact subsets in it have Euclidean metrics and let a measure $\rho_n\in\mathcal{M}(\Omega\times\mathcal{T})$ converges towards a measure $\rho\in\mathcal{M}(\Omega\times\mathcal{T})$ and is uniformly bounded in the variation norm. If the projections of the measure $|\rho_n|$ and $|\rho|$ on $\mathcal{T}$ are uniformly tight and the projections of the measures $|\rho_n|$ on Ω are uniformly countably additive, then*

$$\lim_{n\to\infty}\mathbb{E}_s\int_s^\tau\hat{f}d\rho_n=\mathbb{E}\int_0^T\hat{f}d\rho, \tag{5}$$

where n is the total number of small equal in length subintervals $[s,\tau]$ of $[0,T]$ and the continuous bounded $\mathcal{P}\otimes\mathcal{B}(\mathcal{T})$-Borel measurable function $\hat{f}$ such that,

$$\int_s^\tau\hat{f}d\rho_n=\int_s^\tau\Bigg\{\sum_{i=1}^N\left[Y_i(\nu)-\sum_{j'=1}^J\beta_{j'}(\nu)X_{ij'}(\nu)\right]^2d\nu$$
$$+\lambda[\Delta\mathbf{U}(\nu)-\boldsymbol{\mu}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{d}\nu-\boldsymbol{\sigma}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{dB}(\nu)]\Bigg\},$$

where $\mathcal{P}$ is the probability measure on the Borel σ-algebra $\mathcal{B}(\mathcal{T})$.

Lemma 2. *Suppose, for $\varepsilon\downarrow 0$, $\Psi_{s,s+\varepsilon}$ approximated to a linear function on $\mathcal{F}_s$ within the small time interval $[s,s+\varepsilon]$ such that, Assumptions 1-4, Lemma 1 hold and for $\tilde{f}\geq 0$ we have $\Psi_{s,s+\varepsilon}(\mathbf{X})$, $\lim_{k\to\infty}\Psi_{s,s+\varepsilon}^k(\mathbf{X})\to 0$ for every monotonically decreasing sequence $\tilde{f}_k\in\mathcal{F}_s$. Then there exists a unique measure $N_s^{-1}d\mathbf{X}$ generated by the filtration $\mathcal{F}_s^{\mathbf{X}}$ starting at $\mathbf{X}_0\in\mathbb{R}^{N\times J}$ such that $\mathcal{F}_s^{\mathbf{X}}\subseteq\mathcal{F}_s$ and*

$$\Psi_{s,s+\varepsilon}(\mathbf{X})=\frac{1}{N_s}\int_{\mathbb{R}^{N\times J}}\tilde{f}d\mathbf{X},\ \forall\tilde{f}\in\mathcal{F}_s,$$

where $\tilde{f}=\exp\left[-\varepsilon\mathcal{L}_{s,s+\varepsilon}(\mathbf{X})\right]\Psi_s(\mathbf{X})$.

Proposition 1. *If the objective is to minimize Equation (4) subject to the error dynamics*

$$d\mathbf{U}(s) = \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]ds + \boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]d\mathbf{B}(s), \tag{6}$$

with Assumptions 1-4 and, Lemmas 1, 2, then under continuous time, for $\{i, j\} = \{1, ..., N\}^2$, $j' = 1, ...J$, $X_{ij'}$*'s regression coefficient is found by solving the Equation*

$$2\sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]X_{ij'}(s) - \frac{\partial g[s, \mathbf{X}(s)]}{\partial \mathbf{X}}\frac{\partial \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]}{\partial \boldsymbol{\beta}(s)}\frac{\partial \boldsymbol{\beta}(s)}{\partial \beta_{j'}(s)} - \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial \boldsymbol{\sigma}^{ij}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]}{\partial \boldsymbol{\beta}(s)}\frac{\partial \boldsymbol{\beta}(s)}{\partial \beta_{j'}}\frac{\partial^2 g[s, \mathbf{X}(s)]}{\partial X_{ij'}\partial X_{jj'}} = 0,$$

for $\beta_{j'}$*, with initial condition* $\mathbf{X}_{0_{(N\times J)\times 1}}$*, where* $g[s, \mathbf{X}(s)] \in C^2\left([0, T] \times \mathbb{R}^{N\times J}\right)$ *with* $\mathbf{I}(s) = g[s, \mathbf{X}(s)]$ *is a positive, non-decreasing penalization function vanishing at infinity which substitutes the coefficient dynamics such that,* $\mathbf{I}(s)$ *is an Itô process.*

Example 1. *(LASSO). Consider the dynamic objective function expressed in the Equation (2) subject to the error dynamics*

$$d\mathbf{U}(s) = \sum_{j'=1}^{m}|\beta_{j'}(s)|ds + 2\sum_{i=1}^{N}\sum_{j'=1}^{m}\beta_{j'}(s)X_{ij'}(s)dB(s),$$

where $B(s)$ *is the constant Brownian motion of this system. The main reason of not taking a squared root in the diffusion coefficient is* $\beta_{j'}(s)$ *is small in magnitude. We further assume independent variables evolves exponentially. Therefore, for a positive penalization parameter* λ^**, we assume* $g(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ *where* $\frac{\partial}{\partial X_{ij'}}g(s, X_{ij'}) = sg(s, X_{ij'})$ *and* $\frac{\partial^2}{\partial X_{ij'}^2}g(s, X_{ij'}) = s^2 g(s, X_{ij'})$*. Furthermore, without loss of generality we assume* $m = J$ *and our main concern is to find the optimal coefficient, we assume* $\beta_k \neq 0$ *for any* $k = 1, ..., J$*. Therefore,* $\frac{\partial}{\partial \beta_k}|\beta_k| = \frac{\beta_k}{|\beta_k|}$ *which is* -1 *for all* $\beta_k < 0$ *and* 1 *for all* $\beta_k > 0$*. By using Proposition 1 we have,*

$$2\sum_{i=1}^{N}\left[Y_i(s) - \beta_k(s)X_{ik}(s) - \sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}(s)\right]X_{ij'}(s) - sg(s, X_{ij'})\frac{\beta_k}{|\beta_k|} - s^2\sum_{i=1}^{N}X_{ij'}(s)g(s, X_{ij'}) = 0,$$

which yields,

$$\beta_k = \frac{1}{2\sum_{i=1}^{N}X_{ik}^2(s)}\left\{2\sum_{i=1}^{N}X_{ij'}(s)Y_i(s) - 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}(s) - s\left[g(s, X_{ij'}) + s\sum_{i=1}^{N}X_{ij'}g(s, X_{ij'})\right]\right\},$$

for all $\beta_k > 0$ *and* $\sum_{i=1}^{N}X_{ik}^2(s) \neq 0$ *and,*

$$\beta_k = \frac{1}{2\sum_{i=1}^{N}X_{ik}^2(s)}\left\{2\sum_{i=1}^{N}X_{ij'}(s)Y_i(s) - 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}(s) + s\left[g(s, X_{ij'}) - s\sum_{i=1}^{N}X_{ij'}g(s, X_{ij'})\right]\right\},$$

for all $\beta_k < 0$.

Example 2. *(Ridge regression). Consider again objective function in Equation (2) subject to*

$$d\mathbf{U}(s) = \sum_{j'=1}^{J}\beta_{j'}^2(s)ds + 2\sum_{i=1}^{N}\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)dB(s).$$

Assuming $g(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ *for all* $\sum_{i=1}^N X_{ik}^2 + sg(s, X_{ik}) \neq 0$, *where* $k = 1, ..., J$, *Proposition 1 determines the regression coefficient under ridge regression as*

$$\beta_k = \frac{2\left[\sum_{i=1}^N X_{ij'}(s)Y_i(s) - \sum_{j'=1}^{J-1} \beta_{j'} X_{ij'}^2\right] - s^2 \sum_{i=1}^N X_{ij'}(s) g(s, X_{ij'})}{2\left[\sum_{i=1}^N X_{ik}^2(s) + sg(s, X_{ik})\right]}.$$

Example 3. *(Standard* L^p*-norm). In this framework for all* $p \neq 0$ *let us assume the error dynamics as*

$$d\mathbf{U}(s) = \left[\sum_{j'=1}^J |\beta_{j'}(s)|^p\right]^{\frac{1}{p}} ds + 2\sum_{i=1}^N \sum_{j'=1}^J \beta_{j'}(s) X_{ij'}(s) dB(s),$$

where $dB(s)$ *is the constant Brownian motion in this system such that* $\beta_{j'} \neq 0$ *for all* $j' = 1, ..., J$. *If we minimize the Equation (2) subject to the above coefficient dynamics, Proposition 1 with* $g(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ *gives*

$$2\sum_{i=1}^N Y_i(s) X_{ij'}(s) - 2\sum_{i=1}^N \sum_{j'=1}^J \beta_{j'}(s) X_{ij'}(s) - sg(s, X_{ij'}) \left[\sum_{j'=1}^J |\beta_{j'}(s)|^p\right]^{\frac{1}{p}-1} \sum_{j'=1}^J \beta_{j'} |\beta_{j'}|^{p-2} - s^2 \sum_{i=1}^N X_{ij'} g(s, X_{ij'}) = 0.$$

Hence, for $k = 1, ..., J$ *we have,*

$$2\sum_{i=1}^N Y_i(s) X_{ij'}(s) - 2\beta_k(s) \sum_{i=1}^N X_{ik}^2(s) - 2\sum_{i=1}^N \sum_{j'=1}^{J-1} \beta_{j'}(s) X_{ij'}(s) - sg(s, X_{ik}) \left[|\beta_k(s)|^p\right]^{\frac{1}{p}-1} \beta_k(s) |\beta_k(s)|^{p-2} - sg(s, X_{ij'}) \left[\sum_{j'=1}^{J-1} |\beta_{j'}(s)|^p\right]^{\frac{1}{p}-1} \sum_{j'=1}^{J-1} \beta_{j'}(s) |\beta_{j'}(s)|^{p-2} - s^2 \sum_{i=1}^N X_{ij'} g(s, X_{ij'}) = 0.$$

Furthermore, For all $\beta_k > 0$, $\beta_{j'} > 0$ *and* $\sum_{i=1}^N X_{ik}(s) \neq 0$ *we have,*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)} \left\{ 2\sum_{i=1}^N Y_i(s) X_{ij'}(s) - 2\sum_{i=1}^N \sum_{j'=1}^{J-1} \beta_{j'}(s) X_{ij'}^2(s) - sg(s, X_{ij'}) \left[\sum_{j'=1}^J \beta_{j'}^p(s)\right]^{\frac{1}{p}-1} \sum_{j'=1}^{J-1} \beta_{j'}^{p-1}(s) - sg(s, X_{ik}) - s^2 \sum_{i=1}^N X_{ij'}(s) g(s, X_{ij'}) \right\},$$

and when $\beta_k < 0$ *and* $\beta_{j'} < 0$ *for all* $j', k = 1, ..., J$, *then*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)} \left\{ s^2 \sum_{i=1}^N X_{ij'}(s) g(s, X_{ij'}) - 2\sum_{i=1}^N Y_i(s) X_{ij'}(s) - 2\sum_{i=1}^N \sum_{j'=1}^{J-1} \beta_{j'}(s) X_{ij'}^2(s) - sg(s, X_{ij'}) \left[\sum_{j'=1}^J \beta_{j'}^p(s)\right]^{\frac{1}{p}-1} \sum_{j'=1}^{J-1} \beta_{j'}^{p-1}(s) - sg(s, X_{ik}) \right\}.$$

Example 4. *(Elastic net regression). In this framework for all $\alpha \in [0,1]$ and $\beta_{j'} \neq 0, \forall j' = 1,...,J$ suppose the error dynamics is*

$$d\mathbf{U}(s) = \left\{(1-\alpha)\sum_{j'=1}^{J}|\beta_{j'}(s)| + \alpha\sum_{j'=1}^{J}\beta_{j'}^2(s)\right\}ds + 2\sum_{i=1}^{N}\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)dB(s),$$

where $dB(s)$ is the constant Brownian motion in this system. If we minimize the Equation (2) subject to the above coefficient dynamics, Proposition 1 with $g(s,X_{ij'}) = \lambda^* \exp(sX_{ij'})$ gives

$$2\sum_{i=1}^{N}Y_i(s)X_{ij'}(s) - 2\beta_k(s)\sum_{i=1}^{N}X_{ik}^2(s) - 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s)$$
$$- 2sg(s,X_{ij'})\left[(1-\alpha)\frac{\beta_k(s)}{|\beta_k(s)|} + 2\alpha\beta_k(s)\right] - s^2\sum_{i=1}^{N}X_{ij'}(s)g(s,X_{ij'}) = 0,$$

for $\beta_k > 0$, $\beta_{j'} > 0$ and $\sum_{i=1}^{N}X_{ik}^2 - s\alpha g(s,X_{ij'}) \neq 0$ which gives us

$$\beta_k = \frac{1}{2\left[\sum_{i=1}^{N}X_{ik}^2(s) - s\alpha g(s,X_{ij'})\right]}\Bigg[2\sum_{i=1}^{N}Y_i(s)X_{ij'}(s)$$
$$- 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - s(1-\alpha)g(s,X_{ij'}) - s^2\sum_{i=1}^{N}X_{ij'}(s)g(s,X_{ij'})\Bigg],$$

and, for $\beta_k < 0$, $\beta_{j'} < 0$

$$\beta_k = \frac{1}{2\left[\sum_{i=1}^{N}X_{ik}^2(s) + s\alpha g(s,X_{ij'})\right]}\Bigg[s^2\sum_{i=1}^{N}X_{ij'}(s)g(s,X_{ij'})$$
$$- 2\sum_{i=1}^{N}Y_i(s)X_{ij'}(s) - 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - s(1-\alpha)g(s,X_{ij'})\Bigg].$$

Example 5. *(Fused LASSO). In this framework for all $\alpha \in (0,1)$ let us assume the coefficient dynamics as*

$$d\mathbf{U}(\mathbf{s}) = \left[\alpha\sum_{j'=1}^{J}|\beta_{j'}(s)| + (1-\alpha)\sum_{j'=0}^{J}|\beta_{j'}(s) - \beta_{j'-1}(s)|\right]ds + 2\sum_{i=1}^{N}\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)dB(s),$$

where $dB(s)$ is the constant Brownian motion in this system such that $\beta_{j'} \neq 0$ for all $j' = 1,...,J$. For a function $g(s,X_{ij'}) = \lambda^* \exp(sX_{ij'})$ Proposition 1 yields,

$$2\sum_{i=1}^{N}Y_i(s)X_{ij'}(s) - 2\beta_k(s)\sum_{i=1}^{N}X_{ik}^2(s) - 2\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s)$$
$$- sg(s,X_{ik})\left[\alpha\frac{\beta_k(s)}{|\beta_k(s)|} + (1-\alpha)\frac{\beta_k(s)-\beta_{k-1}(s)}{|\beta_k(s)-\beta_{k-1}(s)|}\right] - s^2\sum_{i=1}^{N}X_{ij'}(s)g(s,X_{ij'}) = 0.$$

Furthermore, for $\sum_{i=1}^{N}X_{ik}^2(s) \neq 0$ if $\beta_k > 0$, $\beta_{j'} > 0$ for all $k = 1,...,J$ such that $\beta_k > \beta_{k-1}$ then

$$\beta_k = \frac{1}{2\sum_{i=1}^{N}X_{ik}^2(s)}\left[2\sum_{i=1}^{N}Y_i(s)X_{ij'}(s) - 2\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - sg(s,X_{ik}) - s^2\sum_{i=1}^{N}X_{ij'}(s)g(s,X_{ij'})\right],$$

if $\beta_k > 0,\ \beta_{j'} > 0$ *such that* $\beta_k < \beta_{k-1}$ *then for* $\sum_{i=1}^N X_{ik}^2 \neq 0$ *we have,*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)}\left[2\sum_{i=1}^N Y_i(s)X_{ij'}(s) - 2\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) + s(1-2\alpha)g(s,X_{ik}) - s^2\sum_{i=1}^N X_{ij'}(s)g(s,X_{ij'})\right],$$

and finally, if $\beta_k < 0,\ \beta_{j'} < 0$ *such that* $\beta_k \neq \beta_{k-1}$ *then*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)}\left[s^2\sum_{i=1}^N X_{ij'}(s)g(s,X_{ij'}) - 2\sum_{i=1}^N Y_i(s)X_{ij'}(s) - 2\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - sg(s,X_{ik})\right].$$

Example 6. *(Bridge regression). For all* $\beta_{j'} \neq 0, \forall j' = 0, ..., J$ *suppose the error dynamics is*

$$d\mathbf{U}(s) = \left\{\sum_{j'=1}^J |\beta_{j'}(s)|^{\frac{1}{2}}\right\}^2 ds + 2\sum_{i=1}^N\sum_{j'=1}^J \beta_{j'}(s)X_{ij'}(s)dB(s),$$

where $dB(s)$ *is the constant Brownian motion in this system. If we minimize the Equation (2) subject to the above coefficient dynamics, Proposition 1 with* $g(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ *gives*

$$2\sum_{i=1}^N Y_i(s)X_{ij'}(s) - 2\sum_{i=1}^N\sum_{j'=1}^J \beta_{j'}(s)X_{ij'}^2(s) - s\beta_{j'}(s)|\beta_{j'}(s)|^{-\frac{3}{2}}g(s,X_{ij'})\sum_{j'=1}^J|\beta_{j'}(s)|^{\frac{1}{2}} - s^2\sum_{i=1}^N X_{ij'}(s)g(s,X_{ij'}) = 0.$$

Furthermore, for all $k = 1, ..., J$, $\beta_k > 0$, $\beta_{j'} > 0$ *and* $\sum_{i=1}^N X_{ik}^2 \neq 0$ *we have,*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)}\Bigg[2\sum_{i=1}^N Y_i(s)X_{ij'}(s) - 2\sum_{i=1}^N\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - sg(s,X_{ik}) - s\beta_{j'}^{-\frac{1}{2}}(s)g(s,X_{ij'})\sum_{j'=1}^{J-1}\beta_{j'}^{\frac{1}{2}}(s) - s^2\sum_{i=1}^N X_{ij'}(s)g(s,X_{ij'})\Bigg],$$

and for $\beta_k < 0$, $\beta_{j'} < 0$ *and* $\sum_{i=1}^N X_{ik}^2 \neq 0$ *we have,*

$$\beta_k = \frac{1}{2\sum_{i=1}^N X_{ik}^2(s)}\Bigg[s^2\sum_{i=1}^N X_{ij'}(s)g(s,X_{ij'}) - 2\sum_{i=1}^N Y_i(s)X_{ij'}(s) - 2\sum_{i=1}^N\sum_{j'=1}^{J-1}\beta_{j'}(s)X_{ij'}^2(s) - sg(s,X_{ik}) - s\beta_{j'}^{-\frac{1}{2}}(s)g(s,X_{ij'})\sum_{j'=1}^{J-1}\beta_{j'}^{\frac{1}{2}}(s)\Bigg].$$

Example 7. *(Group LASSO). For an* m*-dimensional coefficient vector* $\boldsymbol{\beta}_{j'}$ *with* $\mathbf{K}_{j'}$*, an* $m \times m$*-dimensional positive definite matrix assume the error dynamics is,*

$$d\mathbf{U}(s) = \left[\sum_{j'=1}^J \boldsymbol{\beta}_{j'}^T(s)\mathbf{K}_j(s)\boldsymbol{\beta}_{j'}(s)\right]ds + 2\sum_{i=1}^N\sum_{j'=1}^J \boldsymbol{\beta}_{j'}^T\mathbf{X}_{ij'}(s)d\mathbf{B}(s),$$

where $\beta_{j'}^T$ *is the transposition of* $\beta_{j'}$, $\mathbf{X}_{ij'}$ *an* $m \times m$*-dimensional matrix and* $d\mathbf{B}(s)$ *is an* m*-dimensional Brownian motion. Using an* m*-dimensional vector valued function* $g(s, \mathbf{X}_{ij'}) =$

$\lambda^* \exp(s, \mathbf{X}_{ij'})$ and Proposition 1 we get coefficient vector as,

$$\boldsymbol{\beta}_k = \tfrac{1}{2}\left\{\sum_{i=1}^{N} \mathbf{X}_{ik}^T(s)\mathbf{X}_{ik}(s) + \left[\mathbf{K}_k(s) + \mathbf{K}_k^T(s)\right] g(s, \mathbf{X}_{ik})\right\}^{-1}$$
$$\times\Bigg[2\sum_{i=1}^{N} \mathbf{Y}_i(s)\mathbf{X}_{ij'}(s) - 2\sum_{i=1}^{N}\sum_{j'=1}^{J-1} \boldsymbol{\beta}_{j'}(s)\mathbf{X}_{ij'}^T(s)\mathbf{X}_{ij'}(s)$$
$$- sg(s, \mathbf{X}_{ij'})\sum_{j'=1}^{J-1}\left[\mathbf{K}_{j'}(s) + \mathbf{K}_{j'}^T(s)\right]\boldsymbol{\beta}_{j'}(s) - s^2\sum_{i=1}^{N} \mathbf{X}_{ij'}(s) g(s, \mathbf{X}_{ij'})\Bigg],$$

Such that $\left[\sum_{i=1}^N \mathbf{X}_{ik}^T(s)\mathbf{X}_{ik}(s) + \left[\mathbf{K}_k(s) + \mathbf{K}_k^T(s)\right] g(s, \mathbf{X}_{ik})\right]^{-1}$ exists and $\boldsymbol{\beta}_k$ is an m-dimensional vector where $k = 1, ..., J$.

Proposition 2. *Suppose, under the system of smoothing spline regression our objective is to,*

$$\min_{\{\beta_j' \in \boldsymbol{\beta}\}} \overline{\mathbf{X}}_S(s, \boldsymbol{\beta}, \mathbf{X}) = \min_{\{\beta_j' \in \boldsymbol{\beta}\}} \mathbb{E}\int_0^T \sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J} \beta_{j'}(s) h[X_{ij'}(s)]\right]^2 ds, \tag{7}$$

subject to the error dynamics represented by the Equation (1), where h is a dynamic C^2-basis function such that Assumptions 1- 4, Lemmas 1 and 2 hold. Then under continuous time, for $\{i, j\} = \{1, ..., N\}^2$, $j' = 1, ...J$, $h[X_{ij'}]$'s regression coefficient is found by solving the Equation

$$2\sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)h[X_{ij'}(s)]\right] h[X_{ij'}(s)] - \frac{\partial g^*[s, \mathbf{X}(s)]}{\partial \mathbf{X}}\frac{\partial \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]}{\partial \boldsymbol{\beta}(s)}\frac{\partial \boldsymbol{\beta}(s)}{\partial \beta_{j'}(s)}$$
$$- \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial \boldsymbol{\sigma}^{ij}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)]}{\partial \boldsymbol{\beta}(s)}\frac{\partial \boldsymbol{\beta}(s)}{\partial \beta_{j'}}\frac{\partial^2 g[s, \mathbf{X}(s)]}{\partial X_{ij'}\partial X_{jj'}} = 0,$$

for $\beta_{j'}$, with initial condition $\mathbf{X}_{0_{(N\times J)\times 1}}$, where $g^*[s, \mathbf{X}(s)] \in C^2\left([0, T] \times \mathbb{R}^{N\times J}\right)$ with $\mathbf{I}^*(s) = g^*[s, \mathbf{X}(s)]$ is a positive, non-decreasing penalization function vanishing at infinity which substitutes the coefficient dynamics such that, $\mathbf{I}^*(s)$ is an Itô process.

Example 8. *(Cubic smoothing spline) Consider the objective function in Equation (7) where $h[X_{ij'}(s)] = X_{ij'}(s) + X_{ij'}^2(s) + X_{ij'}^3(s)$ subject to the error dynamics*

$$d\mathbf{U}(s) = [2\beta_{j'}(s) + 6\beta_{j'}(s)X_{ij'}(s)]\, ds + 2\sum_{i=1}^{N}\beta_{j'}(s)X_{ij'}(s)dB(s),$$

where $B(s)$ is the Brownian motion under cubic smoothing spline. For a penalization parameter λ^* if we assume $g^*(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ then Proposition 2 gives us,

$$0 = 2\sum_{i=1}^{N}\left\{Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)\left[X_{ij'}(s) + X_{ij'}^2(s) + X_{ij'}^3(s)\right]\right\}$$
$$\times\left[X_{ij'}(s) + X_{ij'}^2(s) + X_{ij'}^3(s)\right] - 6s\beta_{j'}(s)g^*(s, X_{ij'}) - s^2X_{ij'}(s)g^*(s, X_{ij'}),$$

for all $j' = 1, ..., J$. In this case our k^{th} coefficient would be,

$$\beta_k = \frac{1}{2\left\{\sum_{i=1}^N\left[X_{ik}(s) + X_{ik}^2(s) + X_{ik}^3(s)\right]^2\right\}}\Bigg\{2\sum_{i=1}^{N} Y_i(s)\left[X_{ij'}(s) + X_{ij'}^2(s) + X_{ij'}^3(s)\right]$$
$$- 2\sum_{i=1}^{N}\sum_{j'=0}^{J-1}\beta_{j'}(s)\left[X_{ij'}(s) + X_{ij'}^2(s) + X_{ij'}^3(s)\right]^2 - s^2X_{ij'}(s)g(s, X_{ij'})\Bigg\},$$

where $\sum_{i=1}^{N}\left[X_{ik}(s)+X_{ik}^2(s)+X_{ik}^3(s)\right]\neq 0$.

5. Proofs

5.1. Proof of Lemma 1

Without loss of generality assume absolute value of the quantum Lagrangian $|\hat{f}|\leq 1$ and $||\rho_n||\leq 1$, $||\rho||\leq 1$. Suppose, $\gamma_{\mathcal{T}}$ and γ_Ω denote the projections on $\mathcal{T}$ and the sample space Ω, respectively. As $\mathcal{T}$ is a completely regular space, there exists a compact set $\kappa\subset\mathcal{T}$ such that for any $\varepsilon>0$ and for all n we have that,

$$|\rho_n|\circ\gamma_{\mathcal{T}}^{-1}(\mathcal{T}\setminus\kappa)+|\rho|\circ\gamma_{\mathcal{T}}^{-1}(\mathcal{T}\setminus\kappa)\leq\varepsilon.$$

The space $K(\kappa)$ is separable because κ is Euclidean metrizable. For every $\omega\in\Omega$, define a g_ω as a continuous function $s,\mathbf{X}\mapsto\hat{f}(\omega,s,\mathbf{X})$ on κ. Hence, the mapping $g:\Omega\to K(\kappa)$ is Borel. As the projections of measures on Ω are uniformly countably additive, there exists a probability measure θ on $\mathcal{P}$ with respect to which they have uniformly integral densities. By separability of $K(\kappa)$ and applying Lusin's theorem to the mapping g and measure θ, there is a finite partition of Ω into sets $\mathcal{P}_1,...,\mathcal{P}_p,\mathcal{P}_{p+1}\in\mathcal{P}$ and functions $\hat{f}_1,...,\hat{f}_p\in K(\kappa)$ such that $||\mathbb{E}_s\hat{f}_i||_{K(\kappa)}\leq 1$, $||\mathbb{E}_s g_\omega-\mathbb{E}_s\hat{f}_i||_{K(\kappa)}\leq\varepsilon$ for all $\omega\in\mathcal{P}_i$, $i\leq p$, and

$$|\rho_n|\circ\gamma_\Omega^{-1}(\mathcal{P}_{p+1})+|\rho|\circ\gamma_\Omega^{-1}(\mathcal{P}_{p+1})\leq\varepsilon,\forall n.$$

As $\mathcal{T}$ is completely regular, every conditional expectation $\mathbb{E}_s\hat{f}_i$ extends to $\mathcal{T}$ with the preservation of the maximum of the absolute value. By assumption, there exists a time interval index n_0 such that the absolute value of the difference between conditional expected integrals of $\mathbb{E}_s h(\omega,s,\mathbf{X}):=\sum_{i=1}^{p}\mathcal{I}_{\mathcal{P}_i}(\omega)\mathbb{E}_s\hat{f}_i(s,\mathbf{X})$ against the measure ρ_n and ρ does not exceed ε for all $n\geq n_0$, where $\mathcal{I}_{\mathcal{P}_i}(\omega)$ is the indicator function on partition $\mathcal{P}_i$ on $\mathcal{P}$ Bogachev (2007). Furthermore, $\sup_{\mathbf{X}}|\mathbb{E}_s\hat{f}(\omega,s,\mathbf{X})-E_s h(\omega,s,\mathbf{X})|\leq 2$, $|\mathbb{E}_s\hat{f}(\omega,s\mathbf{X})-h(\omega,s,\mathbf{X})|\leq\varepsilon$ on $\cup_{i=1}^{p}\mathcal{P}_i\times\kappa$ with

$$|\rho_n|\left(\Omega\times(\mathcal{T}\setminus\kappa)\right)+|\rho|(\mathcal{P}_{p+1}\times\mathcal{T})\leq\varepsilon.$$

It remains to use the estimate

$$\int_{\Omega\times\mathcal{T}}|\mathbb{E}_s\hat{f}-\mathbb{E}_s h|d|\rho_n|\leq\int_{\cup_{i=1}^{p}\mathcal{P}_i\times\kappa}|\mathbb{E}_s\hat{f}-\mathbb{E}_s h|d|\rho_n|+4\varepsilon\leq 5\varepsilon,$$

and a similar estimate for ρ. Therefore, for $[s,\tau]$ the Equation (5) holds.

5.2. Proof of Lemma 2

(i). Assumption 2 tells us for small time interval $[s,s+\varepsilon]$, $\tilde{f}\leq\tilde{g}$ and $\tilde{f},\tilde{g}\geq 0$ in $\mathcal{F}_s$ two increasing sequences $\{\tilde{f}_{k_1}\}_{k_1\geq 0}$ and $\{\tilde{g}_{k_2}\}_{k_2\geq 0}$ such that $\lim_{k_1\to\infty}\tilde{f}_{k_1}\leq\lim_{k_2\to\infty}\tilde{g}_{k_2}$, hence $\lim_{k_1\to\infty}\Psi_{s,s+\varepsilon}^{k_1}\leq\lim_{k_2\to\infty}\Psi_{s,s+\varepsilon}^{k_2}$, where

$$\Psi_{s,s+\varepsilon}^{k_1}=\frac{1}{N_s^{\tilde{f}_{k_1}}}\int_{\mathbb{R}^{N\times J}}\tilde{f}_{k_1}d\mathbf{X}$$

and,

$$\Psi_{s,s+\varepsilon}^{k_2}=\frac{1}{N_s^{\tilde{g}_{k_2}}}\int_{\mathbb{R}^{N\times J}}\tilde{g}_{k_2}d\mathbf{X}.$$

As $\tilde{f}_{k_1}\leq\lim_{k_2\to\infty}\tilde{g}_{k_2}$, the function $\min(\tilde{f}_{k_1},\tilde{g}_{k_2})\in\mathcal{F}_s$ is increasing to $\tilde{f}_{k_1}$ as $k_2\to\infty$. Which implies,

$$\Psi_{s,s+\varepsilon}^{k_1}=\lim_{k_2\to\infty}\frac{1}{N_s^{\min(\tilde{f}_{k_1},\tilde{g}_{k_2})}}\int_{\mathbb{R}^{N\times J}}\min(\tilde{f}_{k_1},\tilde{g}_{k_2})d\mathbf{X}\leq\lim_{k_2\to\infty}\Psi_{s,s+\varepsilon}^{k_2}.$$

From the above condition we know that, for $\varepsilon \to 0$, the transition function $\Psi_{s,s+\varepsilon} \in \mathcal{F}_s^+$ is independent of the choices of increasing sequences converge in $\mathcal{F}_s^+$ which makes this well defined. Hence, the functionals on $\mathcal{F}_s^+ \cap \mathcal{F}_s$ coincides with initial functionals and conditions 1 and 2 of Assumption 2 hold. If $\tilde{f}_k$ and $\tilde{g}_k$ are non-negative in $\mathcal{F}_s$ and these sequences are increasing to $\tilde{f}$ and $\tilde{g}$ then, we have two monotonic limits as $\max(\tilde{f},\tilde{g}) = \lim_{k\to\infty} \max(\tilde{f}_k, \tilde{g}_k)$ and $\min(\tilde{f},\tilde{g}) = \lim_{k\to\infty} \min(\tilde{f}_k, \tilde{g}_k)$. Condition 3 of Assumption 2 implies $\min(\tilde{f},\tilde{g}) + \max(\tilde{f},\tilde{g}) = \tilde{f} + \tilde{g}$ as we assume $\tilde{f} \leq \tilde{g}$. Now consider, the sequence $\tilde{f}_{k_1,k_2} \geq 0$ defined on $\mathcal{F}_s$ are increasing to $\tilde{f}_{k_2} \in \mathcal{F}_s^+$ as $k_1 \to \infty$. Define $\tilde{g}_{k_3} := \max_{k_2 \leq k_3} \tilde{f}_{k_2,k_3}$ such that $\tilde{g}_{k_3} \in \mathcal{F}_s$. Therefore, as $\tilde{g}_{k_3}$ is an increasing sequence and for each $k_2 \leq k_3$ we have, $\tilde{g}_{k_3} \leq \tilde{g}_{k_3+1}$ and $\tilde{f}_{k_2,k_3} \leq \tilde{g}_{k_3} \leq \tilde{f}_{k_3}$. This implies

$$\frac{1}{N_s^{\tilde{g}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{g}_{k_3} d\mathbf{X} \leq \frac{1}{N_s^{\tilde{g}_{k_3+1}}} \int_{\mathbb{R}^{N\times J}} \tilde{g}_{k_3+1} d\mathbf{X}$$

and,

$$\frac{1}{N_s^{\tilde{f}_{k_2,k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{f}_{k_2,k_3} d\mathbf{X} \leq \frac{1}{N_s^{\tilde{g}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{g}_{k_3} d\mathbf{X} \leq \frac{1}{N_s^{\tilde{f}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{f}_{k_3} d\mathbf{X},$$

as $k_2 \leq k_3$. Hence, $\lim_{k_3\to\infty} \tilde{f}_{k_3} = \lim_{k_3\to\infty} \tilde{g}_{k_3} \in \mathcal{F}_s^+$ and

$$\begin{aligned} \lim_{k_3\to\infty} \frac{1}{N_s^{\tilde{f}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{f}_{k_3} d\mathbf{X} &= \lim_{k_3\to\infty} \frac{1}{N_s^{\tilde{g}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \tilde{g}_{k_3} d\mathbf{X} \\ &= \frac{1}{N_s^{\lim_{k_3\to\infty} \tilde{g}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \left[\lim_{k_3\to\infty} \tilde{g}_{k_3}\right] d\mathbf{X} = \frac{1}{N_s^{\lim_{k_3\to\infty} \tilde{f}_{k_3}}} \int_{\mathbb{R}^{N\times J}} \left[\lim_{k_3\to\infty} \tilde{f}_{k_3}\right] d\mathbf{X}. \end{aligned}$$

Therefore, Condition 4 of Assumption 2 is satisfied.

(ii). Define E as subset of $\mathbb{R}^{N\times J} \times \Omega$ such that the indicator function of this set for $[s, s+\varepsilon]$ is $\mathcal{I}_E \in \mathbb{R}^{N\times J} \times \Omega$ such that any function operating in E is on $\mathcal{F}_s^+$. Now for all $E \subseteq \mathbb{R}^{N\times J} \times \Omega$ set $\mathbf{X}(E) = \Psi_{s,s+\varepsilon}(\mathcal{I}_E)$. As $\mathcal{I}_E \in \mathcal{F}_s^+$, Condition 3 in Assumption 2 holds and for two partitions $E_1, E_2 \subset E$ we have that, $\mathcal{I}_{E_1\cap E_1} = \min(\mathcal{I}_{E_1}, \mathcal{I}_{E_2})$ and $\mathcal{I}_{E_1\cup E_1} = \max(\mathcal{I}_{E_1}, \mathcal{I}_{E_2})$. This implies E is closed with respect to finite unions and intersections. Furthermore, by Condition 4 in Assumption 2 we can say E is closed with respect to countable unions. As we assume $\mathbf{X}$ is a non-negative monotone additive function hence,

$$\mathbf{X}(E_1 \cap E_2) + \mathbf{X}(E_1 \cup E_2) = \mathbf{X}(E_1) + \mathbf{X}(E_2),$$

such that $\mathbf{X} = \lim_{k\to\infty} \mathbf{X}(E_k)$ for all monotonically increasing sequences of sets $E_k \in E$. Hence, there exists a function

$$\mathbf{X}^*(G) = \inf\{\mathbf{X}(E) : \; E \subset \mathbb{R}^{N\times J} \times \Omega, G \subset E\}$$

which is countably measurable on Riemann class,

$$\mathcal{A} = \left\{A \subset \mathbb{R}^{N\times J} : \; \mathbf{X}^*(A) + \mathbf{X}^*(\mathbb{R}^{N\times J} \setminus A) = \mathbf{K}, \; \mathbf{K} \geq 0\right\}$$

and on the Borel class on filtration $\mathcal{F}_s^{\mathbf{X}}$,

$$\mathcal{B} = \{B \subset \Omega : \; \mathbf{X}^*(B) + \mathbf{X}^*(\Omega \setminus B) = 1\}.$$

Define $\mathbf{X}$ as the restriction of $\mathbf{X}^*$ to both $\mathcal{A}$ and $\mathcal{B}$.

(iii). Suppose, a set $A^* \subset \mathcal{A} \times \mathcal{B}$. As $\tilde{f} \in \mathcal{F}_s^+$, then for all constants c we have $\{\tilde{f} > c\} \in \mathbb{R}^{N\times J} \times \Omega$, since

$$\mathcal{I}_{[\tilde{f}>c]} = \lim_{k\to\infty} \min\left[1, k\max(\tilde{f} - c, 0)\right].$$

Therefore, all functions in $\mathcal{F}_s^+$ are measurable with respect to the σ-algebra $\sigma(\mathbb{R}^{N\times J}\times\Omega)$. As we assumed $E\subset\mathbb{R}^{N\times J}\times\Omega$, there exist an increasing sequence of non-negative functions $\tilde{f}_k\in\mathcal{F}_s$ such that $\mathcal{I}_E=\lim_{k\to\infty}\tilde{f}_k$ and $\mathbf{X}^*(E)=\mathbf{X}(E)=\lim_{k\to\infty}\Psi_{s,s+\varepsilon}^k$. Since $\mathbf{X}^*(E)+\mathbf{X}^*(\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E)\geq 1$, it is sufficient to prove that, $\mathbf{X}^*(E)+\mathbf{X}^*(\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E)\leq 1$ to show $E\in\mathcal{A}\times\mathcal{B}$. Hence, it is equivalent to prove

$$\mathbf{X}^*(\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E)\leq\lim_{k\to\infty}\Psi_{s,s+\varepsilon}(1-\tilde{f}_k). \tag{8}$$

As $\tilde{f}_k$'s are increasing sequences, $1-\tilde{f}_k$ are decreasing in $\mathcal{I}_{\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E}$. The positive, finite constant $c\in(0,1)\times\mathbb{R}^{N\times J}$ define a set $\mathcal{E}=\{1-\tilde{f}_k>c\}$ contains $\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E$ and $\mathcal{E}\subset\mathbb{R}^{N\times J}\times\Omega$. Hence, the interval on this new indicator function $\mathcal{I}_{\mathcal{E}}\leq c^{-1}(1-\tilde{f}_k)$ implies

$$\mathbf{X}^*(\{\mathbb{R}^{N\times J}\times\Omega\}\setminus E)\leq\mathbf{X}(\mathcal{E})\leq c^{-1}\Psi_{s,s+\varepsilon}(1-\tilde{f}_k),$$

where constant matrix c^{-1} has each element inverted in it. After keeping the space $\mathbb{R}^{N\times J}$ fixed and letting $c\to 1$ and $k\to\infty$ Inequality (8) is obtained.

(iv). It is important to know that, all the functions in $\mathcal{F}_s^+$ are $\mathcal{F}_s$-measurable. For $E\subset\mathbb{R}^{N\times J}\times\Omega$ if $\tilde{f}=\mathcal{I}_E$, then

$$\Psi_{s,s+\varepsilon}(\mathbf{X})=\frac{1}{N_s}\int_{\mathbb{R}^{N\times J}}\tilde{f}d\mathbf{X} \tag{9}$$

is satisfied by the way $\mathbf{X}$ is defined. Furthermore, Equation (9) holds for any finite linear combinations of indicators of sets in $\mathbb{R}^{N\times J}\times\Omega$. Suppose, a non-negative function $\tilde{f}\in\mathcal{F}_s^+$ and $\tilde{f}\leq 1$. Then for any $k\in\mathbb{N}$, we have that

$$\tilde{f}_k:=\sum_{i=1}^{2^k-1}i2^{-k}\mathcal{I}_{[i2^{-k}<\tilde{f}<(i+1)2^{-k}]}=2^{-k}\sum_{i=1}^{2^{-k}-1}\mathcal{I}_{[\tilde{f}>i2^{-k}]},$$

which follows

$$\Psi_{s,s+\varepsilon}^k(\mathbf{X})=\frac{1}{N_s^{\tilde{f}_k}}\int_{\mathbb{R}^{N\times J}}\tilde{f}_kd\mathbf{X}.$$

From Conditions $1-4$ in Assumption 2 we know as $k\to\infty$, the left and right hand sides of the above equality converges to $\Psi_{s,s+\varepsilon}(\mathbf{X})$ and $\frac{1}{N_s}\int_{\mathbb{R}^{N\times J}}\tilde{f}d\mathbf{X}$ respectively. Moreover, as $\tilde{f}=\lim_{k\to\infty}\min(\tilde{f},k)$ and $\min(\tilde{f},k)\in\mathcal{F}_s^+$ for all $\tilde{f}\geq 0$, Equation (9) still holds. Finally, for any $\tilde{f}\in\mathcal{F}_s$, condition $\tilde{f}=\max(\tilde{f},0)-\max(-\tilde{f},0)$ holds and the uniqueness of $\mathbf{X}$ comes from the fact that E is closed with respect to finite intersections and it generates a σ-algebra.

5.3. Proof of Proposition 1

Using Equations (4) and (6), with initial condition $\mathbf{X}_0$, the Lagrangian of this system is,

$$\mathcal{L}_{0,T}(\mathbf{X})=\int_0^T\mathbb{E}_s\left\{\sum_{i=1}^N\left[Y_i(s)-\sum_{j'=1}^J\beta_{j'}(s)X_{ij'}(s)\right]^2ds+\lambda[\mathbf{U}(s+ds)-\mathbf{U}(s)-\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]ds\right.$$
$$\left.-\boldsymbol{\sigma}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]d\mathbf{B}(s)]\right\},$$

where λ is the time independent Lagrange multiplier which is assumed to be non-negative. Subdivide $[0,T]$ into n equal time-intervals $[s,s+\varepsilon]$. For any positive ε and normalizing constant $N_s>0$, define a transition function as

$$\Psi_{s,s+\varepsilon}(\mathbf{X})=\frac{1}{N_s}\int_{\mathbb{R}^{N\times J}}\exp\left[-\varepsilon\mathcal{L}_{s,s+\varepsilon}(\mathbf{X})\right]\Psi_s(\mathbf{X})d\mathbf{X}, \tag{10}$$

where $\Psi_s(\mathbf{X})$ is the transition function at the beginning of s and $\frac{1}{N_s}\,d\mathbf{X}$ is a finite Riemann measure such that for k^{th} time interval the transition function is,

$$\Psi_{0,T}(\mathbf{X}) = \frac{1}{N_s^n}\int_{\mathbb{R}^{N\times J\times n}} \exp\left[-\varepsilon\sum_{k=1}^{n}\mathcal{L}_{s,s+\varepsilon}^{k}(\mathbf{X})\right]\Psi_0(\mathbf{X})\prod_{k=1}^{n} d\mathbf{X}^k, \tag{11}$$

with the finite measure $N_s^{-n}\prod_{k=1}^{n} d\mathbf{X}^k$ and initial transition function $\Psi_0(\mathbf{X}) > 0$ for all $n \in \mathbb{N}$ Fujiwara (2017). Equations (10) and (11) consider all continuous infinite paths of transition of $\mathbf{X}$ in any two time intervals.

Define $\Delta\mathbf{U}(\nu) = \mathbf{U}(\nu + d\nu) - \mathbf{U}(\nu)$, then Fubuni's theorem implies,

$$\mathcal{L}_{s,\tau}(\mathbf{X}) = \mathbb{E}_s \int_s^{\tau}\left\{\sum_{i=1}^{N}\left[Y_i(\nu) - \sum_{j'=1}^{J}\beta_{j'}(\nu)X_{ij'}(\nu)\right]^2 d\nu \right.$$
$$\left. + \lambda[\Delta\mathbf{U}(\nu) - \boldsymbol{\mu}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{d}\nu - \boldsymbol{\sigma}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{dB}(\nu)]\right\},$$

where $\tau = s + \varepsilon$. As we assume the coefficient dynamics has drift and diffusion parts, $\mathbf{X}(\nu)$ is an Itô process, there exists a smooth function $g[\nu,\mathbf{X}(\nu)] \in C^2([0,T]\times\mathbb{R}^{N\times J})$ such that $\mathbf{I}(\nu) = g[\nu,\mathbf{X}(\nu)]$ where $\mathbf{I}(\nu)$ is an Itô process Øksendal (2003). Assuming

$$g[\nu + \Delta\nu, \mathbf{X}(\nu) + \Delta\mathbf{X}(\nu)] = \lambda[\Delta\mathbf{U}(\nu) - \boldsymbol{\mu}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{d}\nu - \boldsymbol{\sigma}[\nu,\boldsymbol{\beta}(\nu),\mathbf{X}(\nu)]\mathbf{dB}(\nu)],$$

for a very small time interval around s with $\varepsilon \downarrow 0$, generalized Itô's Lemma yields,

$$\begin{aligned}
\varepsilon\mathcal{L}_{s,\tau}(\mathbf{X}) \;=\; & \mathbb{E}_s\left\{\varepsilon\sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]^2 + \varepsilon g[s,\mathbf{X}(s)]\right. \\
& +\varepsilon g_s[s,\mathbf{X}(s)] + \varepsilon g_{\mathbf{X}}[s,\mathbf{X}(s)]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)] \\
& +\varepsilon g_{\mathbf{X}}[s,\mathbf{X}(s)]\boldsymbol{\sigma}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]\Delta\mathbf{B}(s) \\
& \left.+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\varepsilon\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]g_{X_iX_j}[s,\mathbf{X}(s)] + o(\varepsilon)\right\},
\end{aligned}$$

where $\boldsymbol{\sigma}^{ij}[s,\mathbf{X}(s)]$ represents $\{i,j\}^{th}$ component of the variance-covarience matrix, $g_s = \partial g/\partial s$, $g_{\mathbf{X}} = \partial g/\partial\mathbf{X}$ and $g_{X_iX_j} = \partial^2 g/(\partial X_{ij'}\ \partial X_{jj'})$, $\Delta B_i\ \Delta B_j = \delta^{ij}\ \varepsilon$, $\Delta B_i\ \varepsilon = \varepsilon\ \Delta B_i = 0$, and $\Delta X_i(s)\ \Delta X_j(s) = \varepsilon$, where δ^{ij} is the Kronecker delta function. As $\mathbb{E}_s[\Delta\mathbf{B}(s)] = 0$ and $\mathbb{E}_s[o(\varepsilon)]/\varepsilon \to 0$, for $\varepsilon \to 0$, with the vector of initial conditions $\mathbf{X}_{0_{N\times 1}}$ dividing throughout by ε and taking the conditional expectation we get,

$$\begin{aligned}
\mathcal{L}_{s,\tau}(\mathbf{X}) \;=\; & \sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]^2 + g[s,\mathbf{X}(s)] \\
& +g_s[s,\mathbf{X}(s)] + g_{\mathbf{X}}[s,\mathbf{X}(s)]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)] \\
& +\tfrac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{I}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]g_{X_iX_j}[s,\mathbf{X}(s)] + o(1).
\end{aligned}$$

Suppose, there exists a vector $\xi_{N\times 1}$ such that $\mathbf{X}(s)_{N\times 1} = \mathbf{X}(\tau)_{N\times 1} + \xi_{N\times 1}$. For a number $0 < \eta < \infty$ assume $|\xi| \leq \eta\varepsilon[\mathbf{X}^T(s)]^{-1}$, which makes ξ a very small number for each $\varepsilon \downarrow 0$.

Furthermore, as $d\xi$ is a cylindrical measure,

$$\Psi_s^\tau(\mathbf{X}) + \varepsilon\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s} + o(\varepsilon) = \frac{1}{N_s}\int_{\mathbb{R}^{N\times J}}\left[\Psi_s^\tau(\mathbf{X}) + \xi\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial\mathbf{X}} + o(\varepsilon)\right]\times$$
$$\exp\left\{-\varepsilon\left[\sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)[X_{ij'}(\tau)+\xi]\right]^2 + g[s,\mathbf{X}(\tau)+\xi] + g_s[s,\mathbf{X}(\tau)+\xi] + g_{\mathbf{X}}[s,\mathbf{X}(\tau)+\xi]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi] + \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]g_{X_iX_j}[s,\mathbf{X}(\tau)+\xi]\right]\right\}d\xi + o(\varepsilon^{1/2}). \quad (12)$$

After defining a C^2 function

$$\begin{aligned} f[s,\boldsymbol{\beta}(s),\xi] = & \sum_{i=1}^{N}\left[Y_i(s) - \sum_{j'=1}^{J}\beta_{j'}(s)[X_{ij'}(\tau)+\xi]\right]^2 \\ & + g[s,\mathbf{X}(\tau)+\xi] + g_s[s,\mathbf{X}(\tau)+\xi] \\ & + g_{\mathbf{X}}[s,\mathbf{X}(\tau)+\xi]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi] \\ & + \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]\times \\ & g_{X_iX_j}[s,\mathbf{X}(\tau)+\xi], \end{aligned}$$

Equation (12) becomes,

$$\Psi_s^\tau(\mathbf{X}) + \varepsilon\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s} = \frac{1}{N_s}\Psi_s^\tau(\mathbf{X})\int_{\mathbb{R}^{N\times J}}\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\xi]\}\,d\xi + \frac{1}{N_s}\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial\mathbf{X}}\int_{\mathbb{R}^{N\times J}}\xi\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\xi]\}\,d\xi + o(\varepsilon^{1/2}). \quad (13)$$

For $\varepsilon\downarrow 0$, $\Delta\mathbf{X}\downarrow 0$ and

$$f[s,\boldsymbol{\beta}(s),\xi] = f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \sum_{i=1}^{N}f_{X_i}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)][\xi_{ij'} - X_{ij'}(\tau)] + \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}f_{X_iX_j}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)][\xi_{ij'} - X_{ij'}(\tau)][\xi_{jj'} - X_{jj'}(\tau)] + o(\varepsilon). \quad (14)$$

We assume there exists a symmetric, positive definite and non-singular Hessian matrix $\boldsymbol{\Theta}_{(N\times J)\times(N\times J)}$ and a vector $\mathbf{R}_{(N\times J)\times 1}$ such that,

$$\int_{\mathbb{R}^{N\times J}}\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\xi]\}d\xi = \sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\boldsymbol{\Theta}|}}\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\boldsymbol{\Theta}^{-1}\mathbf{R}\}. \quad (15)$$

The second Gaussian integral on the right hand side of Equation (13) becomes,

$$\int_{\mathbb{R}^{N\times J}}\xi\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\xi]\}d\xi = \sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\boldsymbol{\Theta}|}}\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\boldsymbol{\Theta}^{-1}\mathbf{R}\}[\mathbf{X}(\tau) + \tfrac{1}{2}(\boldsymbol{\Theta}^{-1}\,\mathbf{R})]. \quad (16)$$

Equations (14), (15) and (16) imply

$$\Psi_s^\tau(\mathbf{X})+\varepsilon\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s}=\frac{1}{N_s}\sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\mathbf{\Theta}|}}\exp\{-\varepsilon f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]+\tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}$$
$$\times\left\{\Psi_s^\tau(\mathbf{X})+[\mathbf{X}(\tau)+\tfrac{1}{2}(\mathbf{\Theta}^{-1}\mathbf{R})]\frac{\partial\Psi_{\mathbf{s}}^\tau(\mathbf{X})}{\partial\mathbf{X}}\right\}+o(\varepsilon^{1/2}).$$

Assuming $N_s=\sqrt{(2\pi)^{N\times J}/(\varepsilon|\mathbf{\Theta}|)}>0$, we get Wick rotated Schrödinger type equation as,

$$\Psi_s^\tau(\mathbf{X})+\varepsilon\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s}=\{1-\varepsilon f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]+\tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}\times$$
$$\left\{\Psi_s^\tau(\mathbf{X})+[\mathbf{X}(\tau)+\tfrac{1}{2}(\mathbf{\Theta}^{-1}\mathbf{R})]\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial\mathbf{X}}\right\}+o(\varepsilon^{1/2}). \quad (17)$$

For any finite positive number η we know $\mathbf{X}(\tau)\leq\eta\varepsilon|\xi^T|^{-1}$. Then there exists $|\mathbf{\Theta}^{-1}\mathbf{R}|\leq 2\eta\varepsilon|1-\xi^T|^{-1}$ such that for $\varepsilon\downarrow 0$ we have, $\left|\mathbf{X}(\tau)+\frac{1}{2}\ (\mathbf{\Theta}^{-1}\ \mathbf{R})\right|\leq\eta\varepsilon$ and Equation (17) becomes,

$$\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s}=\{-f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]+\tfrac{1}{2}\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}\Psi_s^\tau(\mathbf{X}). \quad (18)$$

As $|\mathbf{\Theta}^{-1}\mathbf{R}|\leq 2\eta\varepsilon|1-\xi^T|^{-1}$, where ξ^T is the transpose of ξ, then at $\varepsilon\downarrow 0$ we can ignore the second term. Therefore, Equation (18) becomes

$$\frac{\partial\Psi_s^\tau(\mathbf{X})}{\partial s}=-f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]\Psi_s^\tau(\mathbf{Z}),$$

and the partial derivative with $\beta_{j'}$ yields,

$$-\frac{\partial}{\partial\beta_{j'}}f[u,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]\Psi_s^\tau(\mathbf{X})=0. \quad (19)$$

In Equation (19) either $\Psi_s^\tau(\mathbf{X})=0$ or $\frac{\partial}{\partial\beta_{j'}}f[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]=0$. As $\Psi_s^\tau(\mathbf{X})$ is a transition wave function it cannot be zero. Therefore, the partial derivative with respect to $\beta_{j'}$ has to be zero. We know, $\mathbf{X}(\tau)=\mathbf{X}(s)-\xi$ and for $\xi\downarrow 0$ as we are looking for some stable solution therefore, in Equation (19) $\mathbf{X}(\tau)$ can be replaced by $\mathbf{X}(s)$. Hence,

$$f[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]=\sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]^2+g[s,\mathbf{X}(s)]+g_s[s,\mathbf{X}(s)]+g_{\mathbf{X}}[s,\mathbf{X}(s)]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]$$
$$+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]g_{X_iX_j}[s,\mathbf{X}(s)]. \quad (20)$$

Equations (19) and (20) then imply

$$2\sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)X_{ij'}(s)\right]X_{ij'}(s)-g_{\mathbf{X}}[s,\mathbf{X}(s)]\frac{\partial\boldsymbol{\mu}[u,\boldsymbol{\beta}(s),\mathbf{X}(s)]}{\partial\boldsymbol{\beta}(s)}\frac{\partial\boldsymbol{\beta}(s)}{\partial\beta_{j'}(s)}$$
$$-\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}g_{X_iX_j}[s,\mathbf{X}(s)]\frac{\partial\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]}{\partial\boldsymbol{\beta}(s)}\frac{\partial\boldsymbol{\beta}(s)}{\partial\beta_{j'}(s)}=0. \quad (21)$$

Optimal $\beta_{j'}(s)$ can be obtained by solving Equation (21).

5.4. Proof of Proposition 2

Using Equations (6) and (7), with initial condition $\mathbf{X}_0$ with its basis $h(\mathbf{X}_0)$, the dynamic Lagrangian action of this system of smoothing spline is,

$$\mathcal{L}^*_{0,T}(\mathbf{X}) = \int_0^T \mathbb{E}_s \left\{ \sum_{i=1}^N \left[Y_i(s) - \sum_{j'=1}^J \beta_{j'}(s) h[X_{ij'}(s)] \right]^2 ds + \lambda^* [\mathbf{U}(s+ds) - \mathbf{U}(s) - \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] ds - \boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] d\mathbf{B}(s)] \right\},$$

where λ^* is the time independent non-negative penalizing constant. After subdividing $[0,T]$ into n equal time-intervals $[s, s+\varepsilon]$ such that for all ε and $N^*_s > 0$, define a transition function as

$$\Psi^*_{s,s+\varepsilon}(\mathbf{X}) = \frac{1}{N^*_s} \int_{\mathbb{R}^{N\times J}} \exp\left[-\varepsilon \mathcal{L}^*_{s,s+\varepsilon}(\mathbf{X}) \right] \Psi^*_s(\mathbf{X}) d\mathbf{X}, \tag{22}$$

where $\Psi^*_s(\mathbf{X})$ is the transition function at the beginning of s and $\frac{1}{N^*_s}\, d\mathbf{X}$ is a finite Riemann measure such that for k^{th} time interval this transition function is,

$$\Psi^*_{0,T}(\mathbf{X}) = \frac{1}{(N^*_s)^n} \int_{\mathbb{R}^{N\times J\times n}} \exp\left[-\varepsilon \sum_{k=1}^n \mathcal{L}^{k*}_{s,s+\varepsilon}(\mathbf{X}) \right] \Psi^*_0(\mathbf{X}) \prod_{k=1}^n d\mathbf{X}^k, \tag{23}$$

with the finite measure $(N^*_s)^{-n} \prod_{k=1}^n d\mathbf{X}^k$ and initial transition function $\Psi^*_0(\mathbf{X}) > 0$ for all $n \in \mathbb{N}$. Equations (22) and (23) consider all continuous infinite paths of transition of $\mathbf{X}$ in any two time intervals.

Fubuni's theorem implies,

$$\mathcal{L}^*_{s,\tau}(\mathbf{X}) = \mathbb{E}_s \int_s^\tau \left\{ \sum_{i=1}^N \left[Y_i(\nu) - \sum_{j'=1}^J \beta_{j'}(\nu) h[X_{ij'}(\nu)] \right]^2 d\nu + \lambda^* [\Delta \mathbf{U}(\nu) - \boldsymbol{\mu}[\nu, \boldsymbol{\beta}(\nu), \mathbf{X}(\nu)] d\nu - \boldsymbol{\sigma}[\nu, \boldsymbol{\beta}(\nu), \mathbf{X}(\nu)] d\mathbf{B}(\nu)] \right\},$$

where $\tau = s + \varepsilon$ and $\Delta\mathbf{U}(\nu) = \mathbf{U}(\nu + d\nu) - \mathbf{U}(\nu)$. As like before the coefficient dynamics has drift and diffusion parts, $\mathbf{X}(\nu)$ is an Itô process, there exists a smooth function $g^*[\nu, \mathbf{X}(\nu)] \in C^2([0,T] \times \mathbb{R}^{N\times J})$ such that $\mathbf{I}^*(\nu) = g^*[\nu, \mathbf{X}(\nu)]$ where $\mathbf{I}^*(\nu)$ is an Itô process of the smoothing spline. Assuming

$$g^*[\nu + \Delta\nu, \mathbf{X}(\nu) + \Delta\mathbf{X}(\nu)] = \lambda^* [\Delta\mathbf{U}(\nu) - \boldsymbol{\mu}[\nu, \boldsymbol{\beta}(\nu), \mathbf{X}(\nu)] d\nu - \boldsymbol{\sigma}[\nu, \boldsymbol{\beta}(\nu), \mathbf{X}(\nu)] d\mathbf{B}(\nu)],$$

for a very small time interval around s with $\varepsilon \downarrow 0$, generalized Itô's Lemma yields,

$$\begin{aligned} \varepsilon \mathcal{L}^*_{s,\tau}(\mathbf{X}) = \mathbb{E}_s \Bigg\{ & \varepsilon \sum_{i=1}^N \left[Y_i(s) - \sum_{j'=1}^J \beta_{j'}(s) h[X_{ij'}(s)] \right]^2 + \varepsilon g^*[s, \mathbf{X}(s)] \\ & + \varepsilon g^*_s[s, \mathbf{X}(s)] + \varepsilon g^*_{\mathbf{X}}[s, \mathbf{X}(s)] \boldsymbol{\mu}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] \\ & + \varepsilon g^*_{\mathbf{X}}[s, \mathbf{X}(s)] \boldsymbol{\sigma}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] \Delta\mathbf{B}(s) \\ & + \tfrac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \varepsilon \boldsymbol{\sigma}^{ij}[s, \boldsymbol{\beta}(s), \mathbf{X}(s)] g^*_{X_i X_j}[s, \mathbf{X}(s)] + o(\varepsilon) \Bigg\}, \end{aligned}$$

where $\boldsymbol{\sigma}^{ij}[s,\mathbf{X}(s)]$ represents $\{i,j\}^{th}$ component of the variance-covarience matrix, $g_s^*=\partial g^*/\partial s$, $g_{\mathbf{X}}^*=\partial g^*/\partial \mathbf{X}$ and $g_{X_iX_j}^*=\partial^2 g^*/(\partial X_{ij'}\ \partial X_{jj'})$, $\Delta B_i\ \Delta B_j=\delta^{ij}\ \varepsilon$, $\Delta B_i\ \varepsilon=\varepsilon\ \Delta B_i=0$, and $\Delta X_i(s)\ \Delta X_j(s)=\varepsilon$, where δ^{ij} is the Kronecker delta function. As $\mathbb{E}_s[\Delta\mathbf{B}(s)]=0$ and $\mathbb{E}_s[o(\varepsilon)]/\varepsilon\to 0$, for $\varepsilon\downarrow 0$, with the vector of initial conditions $\mathbf{X}_{0_{N\times 1}}$ dividing throughout by ε and taking the conditional expectation we get,

$$\begin{aligned}\mathcal{L}_{s,\tau}^*(\mathbf{X}) &= \sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)h[X_{ij'}(s)]\right]^2+g^*[s,\mathbf{X}(s)]\\ &\quad+g_s^*[s,\mathbf{X}(s)]+g_{\mathbf{X}}^*[s,\mathbf{X}(s)]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]\\ &\quad+\tfrac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{I}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]g_{X_iX_j}^*[s,\mathbf{X}(s)]+o(1).\end{aligned}$$

Suppose, there exists a vector $\xi_{N\times 1}$ such that $\mathbf{X}(s)_{N\times 1}=\mathbf{X}(\tau)_{N\times 1}+\xi_{N\times 1}$. For a number $0<\eta<\infty$ assume $|\xi|\leq\eta\varepsilon[\mathbf{X}^T(s)]^{-1}$. Furthermore,

$$\begin{aligned}\Psi_s^{\tau*}(\mathbf{X})+\varepsilon\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s}+o(\varepsilon) &= \frac{1}{N_s^*}\int_{\mathbb{R}^{N\times J}}\left[\Psi_s^{\tau*}(\mathbf{X})+\xi\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial\mathbf{X}}+o(\varepsilon)\right]\\ &\quad\times\exp\Bigg\{-\varepsilon\Bigg[\sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)h[X_{ij'}(\tau)+\xi]\right]^2\\ &\quad+g^*[s,\mathbf{X}(\tau)+\xi]+g_s^*[s,\mathbf{X}(\tau)+\xi]\\ &\quad+g_{\mathbf{X}}^*[s,\mathbf{X}(\tau)+\xi]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]\\ &\quad+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]g_{X_iX_j}^*[s,\mathbf{X}(\tau)+\xi]\Bigg]\Bigg\}d\xi+o(\varepsilon^{1/2}).\end{aligned}\tag{24}$$

After defining a C^2 function

$$\begin{aligned}f^*[s,\boldsymbol{\beta}(s),\xi] &= \sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)h[X_{ij'}(\tau)+\xi]\right]^2\\ &\quad+g^*[s,\mathbf{X}(\tau)+\xi]+g_s^*[s,\mathbf{X}(\tau)+\xi]\\ &\quad+g_{\mathbf{X}}^*[s,\mathbf{X}(\tau)+\xi]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]\\ &\quad+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)+\xi]\times\\ &\quad g_{X_iX_j}^*[s,\mathbf{X}(\tau)+\xi],\end{aligned}$$

Equation (24) becomes,

$$\begin{aligned}\Psi_s^{\tau*}(\mathbf{X})+\varepsilon\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s} &= \frac{1}{N_s^*}\Psi_s^{\tau*}(\mathbf{X})\int_{\mathbb{R}^{N\times J}}\exp\left\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\xi]\right\}d\xi\\ &\quad+\frac{1}{N_s^*}\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial\mathbf{X}}\int_{\mathbb{R}^{N\times J}}\xi\exp\left\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\xi]\right\}d\xi+o(\varepsilon^{1/2}).\end{aligned}\tag{25}$$

For $\varepsilon\downarrow 0$, $\Delta\mathbf{X}\downarrow 0$ and

$$\begin{aligned}f^*[s,\boldsymbol{\beta}(s),\xi] &= f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]+\sum_{i=1}^{N}f_{X_i}^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)][\xi_{ij'}-X_{ij'}(\tau)]\\ &\quad+\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}f_{X_iX_j}^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)][\xi_{ij'}-X_{ij'}(\tau)][\xi_{jj'}-X_{jj'}(\tau)]+o(\varepsilon).\end{aligned}\tag{26}$$

We assume there exists a symmetric, positive definite and non-singular Hessian matrix $\mathbf{\Theta}_{(N\times J)\times(N\times J)}$ and a vector $\mathbf{R}_{(N\times J)\times 1}$ such that,

$$\int_{\mathbb{R}^{N\times J}} \exp\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\xi]\}d\xi = \sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\mathbf{\Theta}|}}\exp\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}. \quad (27)$$

The second Gaussian integral on the right hand side of Equation (25) becomes,

$$\int_{\mathbb{R}^{N\times J}} \xi\exp\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\xi]\}d\xi = \sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\mathbf{\Theta}|}}\exp\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}[\mathbf{X}(\tau)+\tfrac{1}{2}(\mathbf{\Theta}^{-1}\ \mathbf{R})]. \quad (28)$$

Equations (26), (27) and (28) imply

$$\Psi_s^{\tau*}(\mathbf{X}) + \varepsilon\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s} = \frac{1}{N_s^*}\sqrt{\frac{(2\pi)^{N\times J}}{\varepsilon|\mathbf{\Theta}|}}\exp\{-\varepsilon f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}\left\{\Psi_s^{\tau*}(\mathbf{X}) + [\mathbf{X}(\tau)+\tfrac{1}{2}(\mathbf{\Theta}^{-1}\mathbf{R})]\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial\mathbf{X}}\right\} + o(\varepsilon^{1/2}).$$

Assuming $N_s^* = \sqrt{(2\pi)^{N\times J}/(\varepsilon|\mathbf{\Theta}|)} > 0$, the Wick rotated Schrödinger type equation is,

$$\Psi_s^{\tau*}(\mathbf{X}) + \varepsilon\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s} = \{1-\varepsilon f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\varepsilon\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}\times \left\{\Psi_s^{\tau*}(\mathbf{X}) + [\mathbf{X}(\tau)+\tfrac{1}{2}(\mathbf{\Theta}^{-1}\mathbf{R})]\frac{\partial\Psi_{\mathbf{s}}^{\tau*}(\mathbf{X})}{\partial\mathbf{X}}\right\} + o(\varepsilon^{1/2}). \quad (29)$$

As $\mathbf{X}(\tau)\leq\eta\varepsilon|\xi^T|^{-1}$, there exists $|\mathbf{\Theta}^{-1}\mathbf{R}|\leq 2\eta\varepsilon|1-\xi^T|^{-1}$ such that for $\varepsilon\downarrow 0$ we have, $\big|\mathbf{X}(\tau)+\frac{1}{2}\ (\mathbf{\Theta}^{-1}\ \mathbf{R})\ \big|\leq\eta\varepsilon$ and Equation (29) becomes,

$$\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s} = \{-f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)] + \tfrac{1}{2}\mathbf{R}^T\mathbf{\Theta}^{-1}\mathbf{R}\}\Psi_s^{\tau*}(\mathbf{X}).$$

As $|\mathbf{\Theta}^{-1}\mathbf{R}|\leq 2\eta\varepsilon|1-\xi^T|^{-1}$, where ξ^T is the transpose of ξ, then we have,

$$\frac{\partial\Psi_s^{\tau*}(\mathbf{X})}{\partial s} = -f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]\Psi_s^{\tau*}(\mathbf{Z}),$$

and the partial derivative with $\beta_{j'}$ yields,

$$-\frac{\partial}{\partial\beta_{j'}}f^*[u,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]\Psi_s^{\tau*}(\mathbf{X}) = 0. \quad (30)$$

In Equation (30) either $\Psi_s^{\tau*}(\mathbf{X})=0$ or $\frac{\partial}{\partial\beta_{j'}}f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(\tau)]=0$. As $\Psi_s^{\tau*}(\mathbf{X})$ is a transition wave function it cannot be zero. Therefore, the partial derivative with respect to $\beta_{j'}$ has to be zero. We know, $\mathbf{X}(\tau)=\mathbf{X}(s)-\xi$ and for $\xi\downarrow 0$ as we are looking for some stable solution therefore, in Equation (30) $\mathbf{X}(\tau)$ can be replaced by $\mathbf{X}(s)$. Hence,

$$f^*[s,\boldsymbol{\beta}(s),\mathbf{X}(s)] = \sum_{i=1}^N\left[Y_i(s)-\sum_{j'=1}^J\beta_{j'}(s)h[X_{ij'}(s)]\right]^2 + g^*[s,\mathbf{X}(s)] + g_s[s,\mathbf{X}(s)] + g_{\mathbf{X}}^*[s,\mathbf{X}(s)]\boldsymbol{\mu}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)] + \tfrac{1}{2}\sum_{i=1}^N\sum_{j=1}^N\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]g_{X_iX_j}^*[s,\mathbf{X}(s)]. \quad (31)$$

Equations (30) and (31) then imply

$$2\sum_{i=1}^{N}\left[Y_i(s)-\sum_{j'=1}^{J}\beta_{j'}(s)h[X_{ij'}(s)]\right]h[X_{ij'}(s)]-g^*_{\mathbf{X}}[s,\mathbf{X}(s)]\frac{\partial\boldsymbol{\mu}[u,\boldsymbol{\beta}(s),\mathbf{X}(s)]}{\partial\boldsymbol{\beta}(s)}\frac{\partial\boldsymbol{\beta}(s)}{\partial\beta_{j'}(s)}$$
$$-\tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}g^*_{X_iX_j}[s,\mathbf{X}(s)]\frac{\partial\boldsymbol{\sigma}^{ij}[s,\boldsymbol{\beta}(s),\mathbf{X}(s)]}{\partial\boldsymbol{\beta}(s)}\frac{\partial\boldsymbol{\beta}(s)}{\partial\beta_{j'}(s)}=0. \quad (32)$$

Optimal $\beta_{j'}(s)$ can be obtained by solving Equation (32).

6. Discussion

In Lemmas 1 and 2 we show the existence of path integral in penalized regression. Proposition 1 helps us determining the coefficients in more generalized LASSO type frameworks. Then we provide seven cases to obtain a closed form β_k, which are functions of $X_{ik}, X_{ij'}, Y_i$ and $\beta_{j'}$. Furthermore, in cases like LASSO, standard L^p-norm, elastic net regression, fused LASSO and bridge regression we assume $\beta_{j'}$'s are non-zero to get rid of the problem of non-differentiability. Proposition 2 determines optimal β coefficients under generalized spline environment where $h(X_{ij'})$ represents any time dependent basis function and Example 8 considers a dynamic cubic smoothing spline. Throughout this paper we assume $g(s, X_{ij'}) = \lambda^* \exp(sX_{ij'})$ and diffusion coefficient as $2\sum_{i=1}^{N}\beta_{j'}X_{ij'}$ to make our result comprehensible and hence, β_k's are easily comparable among our eight examples. In our future research we will extend this idea into more generalized Riemann manifold.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

The authors declare that they did not use any funding to write this paper.

References

Baaquie, B. E. (2007). *Quantum finance: Path integrals and Hamiltonians for options and interest rates*. Cambridge University Press.

Bellman, R. (1966). Dynamic programming. *Science* **153** 34–37.

Bogachev, V. I. (2007). *Measure theory* **2**. Springer Science & Business Media.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical science* **11** 89–121.

Feynman, R. P. (1949). Space-time approach to quantum electrodynamics. *Physical Review* **76** 769.

Fujiwara, D. (2017). *Rigorous time slicing approach to Feynman path integrals*. Springer.

Hall, P. and Opsomer, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92** 105–118.

Hua, L., Polansky, A. and Pramanik, P. (2019). Assessing bivariate tail non-exchangeable dependence. *Statistics & Probability Letters* **155** 108556.

Huang, J. Z. and Su, Y. (2021). Asymptotic Properties of Penalized Spline Estimators in Concave Extended Linear Models: Rates of Convergence. *arXiv preprint arXiv:2105.06367*.

Krylov, N. V. (2008). *Controlled diffusion processes* **14**. Springer Science & Business Media.

Lai, M.-J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica* 1399–1417.

Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95** 415–436.

Øksendal, B. (2003). Stochastic differential equations. In *Stochastic differential equations* 65–84. Springer.

Polansky, A. M. and Pramanik, P. (2021). A motif building process for simulating random networks. *Computational Statistics & Data Analysis* **162** 107263.

Pramanik, P. (2016). *Tail non-exchangeability.* Northern Illinois University.

Pramanik, P. (2020). Optimization of market stochastic dynamics. In *SN Operations Research Forum* **1** 1–17. Springer.

Pramanik, P. (2021a). Effects of water currents on fish migration through a Feynman-type path integral approach under $\sqrt{8/3}$ Liouville-like quantum gravity surfaces. *Theory in Biosciences* **140** 205–223.

Pramanik, P. (2021b). Consensus as a Nash Equilibrium of a stochastic differential game. *arXiv preprint arXiv:2107.05183.*

Pramanik, P. (2021c). Optimization of Dynamic Objective Functions Using Path Integrals, PhD thesis, Northern Illinois University.

Pramanik, P. and Polansky, A. M. (2019). Semicooperation under curved strategy spacetime. *arXiv preprint arXiv:1912.12146.*

Pramanik, P. and Polansky, A. M. (2020a). Optimization of a Dynamic Profit Function using Euclidean Path Integral. *arXiv preprint arXiv:2002.09394.*

Pramanik, P. and Polansky, A. M. (2020b). Motivation to Run in One-Day Cricket. *arXiv preprint arXiv:2001.11099.*

Pramanik, P. and Polansky, A. M. (2021). Scoring a Goal optimally in a Soccer game under Liouville-like quantum gravity action. *arXiv preprint arXiv:2108.00845.*

Prokhorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications* **1** 157–214.

Ramsay, J. O. (2004). Functional data analysis. *Encyclopedia of Statistical Sciences* **4**.

Ramsay, J. O. and Silverman, B. W. (2007). *Applied functional data analysis: methods and case studies.* Springer.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric regression* **12**. Cambridge university press.

Schwarz, K. and Krivobokova, T. (2016). A unified framework for spline estimators. *Biometrika* **103** 121–131.

Wang, X., Shen, J. and Ruppert, D. (2011). On the asymptotics of penalized spline smoothing. *Electronic Journal of Statistics* **5** 1–17.

Yeung, D. W. and Petrosjan, L. A. (2006). *Cooperative stochastic differential games.* Springer Science & Business Media.