LOOKING BACKWARD AND LOOKING FORWARD

ZHENGYUAN GAO

1. Motivation

The equilibrium from rational expectation models bases on the assumption that the agents inside the model assume the model's predictions are valid. As an essential part of the model, the expectation of agents about future economic conditions is still controversial because the expected profit or utility for each agent will be rather different if the assumption of agent's expectation differs. [14] show how agents adjust their long-term expectations under different rules of predictors. [13] introduce adaptive rational equilibrium dynamics, whose predictor is attached to the equilibrium dynamics of certain endogenous variables. Some other long memory predictors have been used by [15]. These analyses, however, rely on the same structure that agents expect the future situation through a single presumably correct law of motion. A recent monograph by [16] introduces the robustness concern to agent's expectation so that the agent's decision contains his/her prior worries about the mis-specification of the expectation. The associated predictor for the robust decision agents is the recursive least square method or the Kalman filter that uses only the first two moments of information of the stochastic motion law. The Kalman filter predictor, as a dual of the linear-quadratic regulator problem of utilities, indicates that agent's prediction only involves parts of observable information. Thus, a class of models whose information is not far from underlying model in certain divergence will be considered when agent makes the expectation. The maximum utilities of robust decision agents are obtained by minimizing the risks towards this class of alternative models.

Although the minimax expectation of robustness concern provides a more general analysis, the alternative models it considers belong to a class of partially specified processes. If we enlarge such a class to an abstract economic system, say an abstract probability space, does the Kalman filter predictor still an optimal choice? An alternative question is that if under certain circumstance the least square type filter is our best choice, then "how far" is it from this circumstance to the reality. The reason of asking these questions is that we realize that the current set-up of rational expectation models is merely an approximation to the real world. The filtering process as agent's predictor in this set-up is to fit the request of the approximating model. We intend to figure out the feasible range of applying filtering as the agent's prediction process. The necessity of doing this is to complement the gap between rational and boundedly rational literature in the stochastic dynamical system. Rational models, where agents face restrictive alternative models, may give us similar interpretations as the boundedly rational models do, where agents use restrictive information. The paper also intends to mitigate the debates in econometrics of using frequentist or Bayesian approaches for estimating dynamic economic system. The Bayesian method comes out as a consequence of several restrictions on the abstract economic model while frequentist method stays away from parts of these restrictions. A Bayesian estimation may be merely a compromise to the real complex world while frequentist estimation has to suffer the intractability from the complexity.

The paper aims at making previous intuitive descriptions rigorous. The mathematical tools we use here are from stochastic analysis and stochastic control. In section 2, we will prove the existence of the equilibrium conditional distribution process for expectation given that fact that the underlying model will generate unpredictable events and observable unpredictable events will affect the expectation process. Section 3 considers an approximation of the abstract economy. The approximation shows that a tractable problem of filtering relies on several assumptions about rationality, fairness, and risk neutral. Section 4 provides a specific representation for agent's expectation. The representation induces a computational demanding problem of filtering. Gaussian prior process, stochastic approximation and extrapolation have been use to rescue this computational infeasible problem. A simulated example of complex model is given in section 5.

2. Equilibrium

2.1. **A model with $\mathbb{P}$-null set information.** The economic system in this paper is driven by some intrinsic elements whose evolutions are modelled via certain evolving processes. We stack these intrinsic elements into a *state vector* and denote it as $X_t = \{X_{i,t}, t > 0, i = 1, \dots\}$. When we state specifically $t \in \mathbb{Z}$, X_t is a discrete time process, otherwise X_t is assumed as a continuous time process. In the paper, x refers to either a deterministic variable or a realization of X_t. If X_t consists of unobservable features such as private information, utilities or underlying prices, then X_t includes hidden states. The evolution of X will be considered as affect some an *observable process* $Y_t = \{Y_{i,t}, t > 0, i = 1, \dots\}$ as observable choices. A correspondence should exist between states and observations.

The *underlying abstract economic model* in our context is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where we define X together with a filtration $(\mathcal{F}_t)_{t\geq 0}$. In $\mathcal{F}$, there is a $\mathbb{P}$-null set[1]containing in $\mathcal{F}_0$ and consequently in all $\mathcal{F}_t$. The algebra $\mathcal{F}$ is $\lim_{t\to\infty} \mathcal{F}_t$. The filtration $\mathcal{F}_t$ is right continuous, $\mathcal{F}_t := \cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}$.

[1]A set A is called $\mathbb{P}$-null set if A is measurable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}(A) = 0$.

The values of states X_t form a measurable space $(\mathbb{S}, \mathcal{S})$. The state space $\mathbb{S}$ is a compact metric space and associates with Borel σ-algebra $\mathcal{S} = \mathcal{B}(\mathbb{S})$. We assume X to be measurable and the measurable mapping is:

$$X_t : ([0,\infty) \times \Omega,\, \mathcal{B}([0,\infty)) \otimes \mathcal{F}) \to (\mathbb{S}, \mathcal{S}),$$

where $\otimes$ denotes product σ-field.

While the essential features of economic dynamics are assumed to be captured by the state variables X_t, the observable economic variables or those of public interest may not be directly included in X_t. The observable and public information is the major resource for agents to make their expectations about how the economic states X_t change. Thus, we need to specify this kind of information. We let Y_t include these observable variables that are relating with X_t. The *observable information set* is $\mathcal{Y} := \vee_{t\in\mathbb{R}^+}\mathcal{Y}_t$ the filtration generated by the observable process Y_t

$$\mathcal{Y}_t := \sigma(Y_s, s \in [0,t]) \vee \mathcal{N}$$

with $t \geq 0$, where $\mathcal{N}$ is the collection of all $\mathbb{P}$-null sets of our economic model $(\Omega, \mathcal{F}, \mathbb{P})$. Notation $A \vee B$ means that the σ-algebra is generated by A and B. The σ-algebra $\mathcal{Y}_t$ is the available information induced by observations up to time t and thus it will be used for making inferences of X.

Agents in model $(\Omega, \mathcal{F}, \mathbb{P})$ will make the inference of states X_t via $\mathcal{Y}_t$. It means computing or approximating some quantities of X_t in terms of a conditional expectation, in other words, computing π_t, the *conditional distribution* of X_t, given $\mathcal{Y}_t$. For any t, the conditional distribution is a stochastic process $(\omega, t) \mapsto \pi_t^\omega$ such that

$$\pi_t^\omega(A) = \mathbb{P}\left[X_t \in A | \mathcal{Y}_t\right](\omega), \qquad A \subset \mathcal{S}.$$

For simplicity, we write π_t^ω as π_t in short. Furthermore, the conditional expectation of X_t is an equivalence class of $\mathcal{Y}_t$-measurable Xs:

$$\mathbb{P}[X_t \cap B] = \mathbb{P}[X \cap B], \qquad X, B \subset \mathcal{Y}_t.$$

Because of $\mathcal{N} \subset \mathcal{Y}_t$, π_t may not be well defined for all $\omega \in \Omega$ but only for ω outside the $\mathbb{P}$-null set. Thus the question of existence of π is equivalent to the question that under what circumstances one can gain sufficient control over all $\mathbb{P}$-null sets $\mathcal{N}$ so that the integration $\int \varphi(x)\pi_t(dx)$ makes sense for a certain class of *choice functions* φ.

Note that X is $\mathcal{F}_t$-adapted but expectation is evaluated conditioning on $\mathcal{Y}_t$ in reality instead of $\mathcal{F}_t$ in the underlying abstract economy $(\Omega, \mathcal{F}, \mathbb{P})$. It implies that even if some relations between X_t and Y_t exist, for example Y_t partially depends on X_t, but expectations conditional on $\mathcal{Y}_t$ does *not* necessarily coincide with those conditional on $\mathcal{F}_t$. Especially, $\mathbb{P}$-null information in $\mathcal{Y}_t$ implies that the unpredictable events will affect the expectation once these events have been observed by the agents. In next section, we will show that an equilibrium conditional density exists even the information set and (mechanism of economic state) do not "match".

2.2. **Existence of equilibrium.** We have assumed that $\mathcal{F}_t$ is right continuous. To make Y_t comparable with X_t, we assume that the filtration $\mathcal{Y}_t$ is also *right continuous*. Let $\mathcal{P}(\mathbb{S})$ denote the space of all probability measures on $\mathbb{S}$. Let ${}^o\varphi(X_t)$ be a process that is defined on the smallest σ-algebra on $([0,\infty)\times\Omega, \mathcal{B}([0,\infty))\otimes\mathcal{F})$ such that every $\mathcal{Y}_t$-adapted process is measurable. Then ${}^o\varphi(\cdot)$ is a $\mathcal{Y}$-measurable function of stochastic states X_t. Moreover, ${}^o\varphi(\cdot)$ can be thought as the $\mathcal{Y}$-measurable representation of the expected choice $\int\varphi(x)\pi_t(dx)$. Therefore, the existence of ${}^o\varphi(\cdot)$ induces the existence of $\pi_t(\cdot)$ and vice versa.

Theorem 1. *For a compact set $\mathbb{S}$ and its Borel σ-algebra $\mathcal{S}$, there is a $\mathcal{P}(\mathbb{S})$-valued conditional distribution process π_t such that for any bounded $\mathcal{S}$-measurable function $\varphi\in B(\mathbb{S})$,*

$$\mathbb{P}\left[\int_{\mathbb{S}}\varphi(x)\pi_t(dx) = {}^o\varphi(X_t) \quad \forall t\right] = 1.$$

Proof. In appendix. □

The process π_t is $\mathcal{Y}_t$-adapted while X_t is $\mathcal{F}_t$-adapted, thus $\mathbb{E}[\varphi(X_t)|\mathcal{Y}_t]$ maybe not well-defined. With an enlarged σ-algebra, a representative process will be defined for $\varphi(X_t)$ even if X is not $\mathcal{Y}_t$-adapted [5, theorem 7.1]. This theorem is called projection theorem. The theorem says that if a process X is measurable and bounded, then for every stopping time T, there is a representation ${}^oX$ (optional process) such that

$$ {}^oX_T\mathbb{I}_{\{T<\infty\}} = \mathbb{E}[X_T\mathbb{I}_{\{T<\infty\}}|\mathcal{Y}_T], \tag{2.1}$$

where $\mathbb{I}_{\{A\}}$ is an indicator function for a set A. Note that there is no restriction on the stopping time T.

Remark. Mathematically, the function φ is also called test function in distribution theory and the process ${}^o\varphi(X_t)$ is called optional process in stochastic analysis. The reason of introducing an optional process is to fill in the gap between the *observable information set* $\mathcal{Y}_t$ and the model's filtration $\mathcal{F}_t$. The procedure of finding an optional process is often called projection. The purpose of this projection is to get an approximating representation on the available information set.[2] It is a way of *looking backward.* The constructed conditional distribution is used by the agents to make expectations. So it is a way of *looking forward.*

Remark. The proof is an analytic procedure of proving the existence of a positive functional measure (a random probability measure). Step 1-3 is modified from theorem 5.1.15 in [2].

[2]In the next section, we will give an explicit expression of this projection by defining a generator. In economic models, this generator can be used to describe agent's (local) maximal principle.

Instead of using the complete separable $\mathbb{S}$, we prefer the compact metric space $\mathbb{S}$ here. There is a subtle difference between these conditions, but since a complete separable space is homeomorphic to a subset of a compact metric space, i.e. theorem 6.6.40 [7], in the following, the subtlety between complete separable space and compact metric space of $\mathbb{S}$ is often ignored on purpose. But for this theorem, the compactness allows the finite sum of φ_i so that the mapping between φ_i and g_i preserves the linearity and furthermore the Riesz representation is more easily to apply on a compact set.

Remark. The uniqueness result in [2] does not hold. Compactness does regularizes the $\mathbb{P}$-null set of $\mathbb{S}$ but not the $\mathbb{P}$-null set in the observable information set $\mathcal{Y}_t$. Thus although we can construct a π_t^ω conditional on $\mathcal{Y}_t$, without any additional regularization of $\mathcal{Y}_t$, the uniqueness of π_t^ω is not verified.

Although it complicates the set-up, $\mathbb{P}$-null set is the crucial feature in this section. Apart from its mathematics characteristics, it has fruitful meanings in economic problems and motivates us to ponder over the manner of judging a model by statistical data.

The role of $\mathbb{P}$-null set in defining a conditional probability is first discovered and illustrated by Kolmogorov in his famous *Borel–Kolmogorov paradox*. The paradox shows that the conditional probability is not uniquely defined with respect to a null set, see Chapter 5 [4]. From the economic perspective, one can think the $\mathbb{P}$-null set on $\mathcal{F}$ and $\mathcal{Y}$ as those unexpected and rare events which have been respectively included in the underlying economic mechanism $\mathcal{F}$ and in the agent's observable information set $\mathcal{Y}$ but not in $\mathcal{F}$. To see the difference, let's assume the $\mathbb{P}$-null events in $\mathcal{Y}$ as consequences of aggregating of those countable $\mathbb{P}$-null events in $\mathcal{F}$ and thus they are too "complex" to be embedded in the underlying model, the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The model $(\Omega, \mathcal{F}, \mathbb{P})$ attributes zero-measures for any countable events that go beyond its capacity, but for uncountable events the model does not even has ability to interpret their existences. For $\mathbb{P}$-null set in $\mathcal{Y}$, individuals may set arbitrary beliefs towards the events on the set, because they cannot figure out any "law" on the set. More discussions about how to specify the law will be given in section 3. In the rest of this section, we will discuss the economic meaning of $\mathbb{P}$-null set on $\mathcal{Y}$ which we will call *overflow* and we will see how it relates to the regularized $\mathbb{P}$-null set on $\mathcal{F}$, a consequence of theorem 1.

Overflow may conflict some common senses and is not a pleasant gift for modelling, ad-hoc testing and forecasting. Before giving a quantitative argument about its importance, we use a qualitative example to discuss this phenomenon. We interpret economic bubbles by using the concept of overflow on the$\mathbb{P}$-null set.

The debate that whether or not an economic bubble exists has a long history among the economists. Here our intention is to use the bubble as an example to explain the *overflow* characteristic instead of joining the debate. Suppose individual gamblers observe arbitrage opportunities in the hedging and the public realize the gains of the gamblers. The speculations, therefore, are included in the observable information set $\mathcal{Y}_t$ for the agents. However, the strategies behind these speculations are untouchable for the public or the society and are conducted thorough inhomogeneous ways, such as forbidden disclosures (private information), special equipments (high-frequency trading), or even improper policies (lobbying), etc.. Any economic model that wants to cover some or all these specific features will make its complexity blow up and this limitation is recognized by the public and the society, thus it is reasonable for the public or the society to believe that the underlying economic model $(\Omega, \mathcal{F}, \mathbb{P})$ will set zero measure on each of these strategies and the associated actions. In other words each action of the speculation is in the $\mathbb{P}$-null set on $(\Omega, \mathcal{F}, \mathbb{P})$. The economic bubble can be considered as an aggregated effect of these actions. Since there are numerous speculations happening in every minute, it is natural to think that the collection of these actions is uncountable. Later, we will show that an uncountable collection of null set is not necessarily contained in the $\mathbb{P}$-null set. It means that the aggregated effect, the bubble, may have a positive probability to occur, namely appear in $\mathcal{Y}$. And in fact, it does. So one should realize the probability model is not a "proper" model but a model that "compromises" to unknowns.

To formalize the previous argument, let $A_1,\, A_2, \ldots \in \mathcal{S}$ be a sequence of pairwise disjoint sets. In order to ensure that π_t is a regular conditional distribution, the σ-additivity condition needs to be satisfied:

$$\pi_t^{\omega}\left(\cup_i A_i\right) = \sum_{i=1} \pi_t^{\omega}(A_i)$$

for every $\omega \in \Omega \backslash \mathcal{N}(A_i,\, i \geq 1)$. The $\mathcal{N}(A_i,\, i \geq 1)$ is the $\mathbb{P}$-null set for the disjoint set A_i with any $i \geq 1$. Let the collection of these null sets be $\mathcal{N}$. Note that the power set of all null set $\{A_i\}$ is $2^{\mathbb{N}}$ which is uncountable. It means that $\mathcal{N}$ is uncountable. We know that π_t^{ω} satisfies the σ-additivity condition only if $\omega \in \mathcal{N}(A_i)$ for any $i \geq 1$ but not $\omega \in \mathcal{N}$. Therefore, some event in $\mathcal{N} \backslash \mathcal{N}(A_i,\, i \geq 1)$ is not in the null sets and has positive probability to occur

$$\mathcal{Y} \cap \{\mathcal{N} \backslash \mathcal{N}(A_i,\, i \geq 1)\} \neq \varnothing.$$

In fact, the set $\mathcal{N}$ need not even be measurable because it is defined in terms of an uncountable union. Then π_t^{ω} cannot be a probability measure. The purpose of theorem 1 is to regularize this problem so that the projected π_i is on a countable subspace. This regularization implicitly forces π_t^{ω} to ignore those aggregated effects or the collections of countable $\mathbb{P}$-null sets on $\mathcal{F}$.

Beside the overflow, the other effect is the arbitrary definition of π_t^{ω} over the $\mathbb{P}$-null set on $\mathcal{Y}$. The arbitrariness allows us to modify $\mathcal{Y}_t$-adapted processes by changing the values of these processes (change of measure) on the $\mathbb{P}$-null set and then the new process should be still $\mathcal{Y}_t$-adapted. The new process is helpful for simplifying the analysis but it induces an arbitrary distribution class for π_t^ω. Therefore in order to construct a satisfied approximating model, additional regularization about the null set are required. Next section will concern this issue.

3. An Approximating Econom(etr)ic Model

In previous section, a general model shows that an equilibrium conditional density process π_t^ω exists even some observable (negligible) events in the system are not caused by the equilibrium mechanism. It also shows that if the model needs a regular solution, it should, by some means, get rid of these irregular events. On the other hand, the general model does not infer any explicit solution because the process ${}^o\varphi(X)$ is somehow built on the top of a "cloud". In this section, we look for an approximating model that will regularize the abstract process ${}^o\varphi(X)$ and exploit a specific approximation of ${}^o\varphi(X)$.

There are three claims in the section: 1. to regularize a class of probabilities that are not uniquely defined on the $\mathbb{P}$-null set on $\mathcal{F}$, 2. to construct an approximating model of X that is embedded in the general model $(\Omega, \mathcal{F}, \mathbb{P})$, 3. to specify the motions of observable process Y. Claim 1 and 2 basically concern the same issue of finding a feasible sub-class model of X but the development of claim 2 depends on claim 1. Claim 3. concerns the issue of constructing a feasible representation of ${}^o\varphi(X_t)$ based on observable process Y.

3.1. **Fairness Existence.** The following claim gives us a "stochastic constant" upon where we can build our model:

Claim. (Martingale Fairness[3] or Contingent claim, MF) A probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ is absolutely continuous with respect to $\mathbb{P}$, such that $\mathbb{Q} \sim \mathbb{P}$. The information of state X_t at any time t is "fair" for all agents under $\mathbb{Q}$ (and has Markov structure[4]).

The "fair" condition means the *martingale* property of X:

$$\mathbb{E}_{\mathbb{Q}}[X_t | \mathcal{F}_s, s \leq t] = X_s \quad \text{and} \quad \mathbb{E}_{\mathbb{Q}}[X_t - X_s | \mathcal{F}_s, s \leq t] = 0.$$

To simply illustrate this condition, suppose any adjustment over the state X at time T has a value H_T and H_T is predictable i.e. each H_T is $\mathcal{F}_t$-measurable for $t < T$. Let the state's pay-off at time t be a random variable $\tilde{f} = \int_0^T H_t dX_t$ where the integral is the Itô integral[5]. The "fairness" says that any $\tilde{f}$ constructed in this way will have zero expected pay-off such that $\mathbb{E}_{\mathbb{Q}}[\tilde{f}] = 0$.

[3]The problem can be extended to a semi-martingale problem by using No Free Lunch claim (Kreps-Yan Theorem). But then X in general cannot provide any explicit solution form for the conditional probability π.

[4]$\mathbb{E}[f(X_T)|\mathcal{F}_t] = \mathbb{E}[f(X_T)|X_t]$ for any $f(\cdot) \in B(\mathbb{S})$.

[5]The definition of Itô integral is given in the Appendix.

The martingale model $(\Omega, \mathcal{F}, \mathbb{Q})$ is treated as a *ghost* model since it may never happen in the reality. However, if one accepts the existence of this martingale model, it will guide us to a feasible model and help us to solve it. With absolute continuity of $\mathbb{Q}$ and $\mathbb{P}$, if there is a $\mathbb{P}$-martingale process Z on $(\Omega, \mathcal{F})$, then any $\mathbb{Q}$-martingale process X implies an $\mathbb{P}$-martingale process ZX. It is obvious that a process is regularized on either measure then it will be regularized on the other one.

It is better to consider the martingale property together with the Markov structure of X such that the filtration $\mathcal{F}_s$ is independent of the $\mathcal{F}$-adapted X_u if $s < t < u$. For arbitrary time $t < u$, there is a transition kernel $\mathbb{Q}_{u-t}(X_u|X_t)$. Chapman-Kolmogorov equation says that

$$\mathbb{Q}_{u-s}(X_u|X_s) = \int \mathbb{Q}_{u-t}(X_u|X_t)\mathbb{Q}_{t-s}(dX_t|X_s)$$

which can be simply stated as $\mathbb{Q}_{\tau+\tau'}(\cdot|\cdot) = \mathbb{Q}_{\tau'}\mathbb{Q}_{\tau}$ for $\tau = t - s$, $\tau' = u - t$. The existence of the kernel $\mathbb{Q}_{\tau'}(\cdot|\cdot)$ is a direct result of the Kolmogorov existence theorem [11, theorem 7.4]. When the process is assumed to be homogeneous on time, the family of $\mathbb{Q}(\cdot|\cdot)$ is a semigroup of transition kernels and has been extensively studied in the recent works of operator methods, see e.g. [12]. It is obvious that the transition kernel $\mathbb{Q}_{\tau'}(\cdot|\cdot)$ is a regular condition probability.

Since the evolution of the state is completely captured by $\mathbb{Q}_{\tau'}(\cdot|\cdot)$, the variation of this transition kernel describes the variation of the evolution pattern of X. It extracts important characteristics in the dynamics. To specify this element, we need the tools of expansion. Take the transition probability $\mathbb{Q}_{\tau'}$ and expand it w.r.t. τ' at zero by Taylor's expansion:

$$\mathbb{Q}_{\tau'}(X_u|X_t) = \delta(X_u - X_t) + \tau'\mathcal{W}(X_u|X_t) + o(\tau'), \tag{3.1}$$

where $\delta(\cdot)$ is the delta function[6]. The function $\mathcal{W}(X_u|X_t)$ is the time derivative of the transition probability at $\tau = 0$, called *transition probability per unit time*. This expression must satisfy the normalization property, in other words, the integral over X_u must equal one. For this purpose, the above form can be corrected to:

$$\mathbb{Q}_{\tau'}(X_u|X_t) = (1 - \alpha_0\tau')\delta(X_u - X_t) + \tau'\mathcal{W}(X_u|X_t) + o(\tau'),$$

where $\alpha_0(X_u) = \int \mathcal{W}(X_u|dX_t)$. Substituting the expansion form into Chapman-Kolmogorov equation, dividing the equation by τ' and then letting τ' go to zero give us the following result.

[6]Loosely speaking, delta function is a smooth indicator function such that the derivative of $\delta(\cdot)$ exists in the weak sense. Regardless of technical differences, one can think both of them are identical.

Corollary 2. *The martingale model* $(\Omega, \mathcal{F}, \mathbb{Q})$ *implies a gain-loss equation for the system such that:*

$$\frac{\partial}{\partial \tau}\mathbb{Q}_\tau(X_u|X_s) = \int \left\{ \mathcal{W}(X_u|X_t)\mathbb{Q}_\tau(dX_t|X_s) - \mathcal{W}(dX_t|X_u)\mathbb{Q}_\tau(X_u|X_s) \right\}.$$

The first term is the gain of state X_u *due to transition from other states* X_t *and the second term is the loss due to transitions from* X_u *into other states.*

Proof. For details, please see Chapter 5 in [6]. □

If the $\partial_\tau \mathbb{Q}_\tau(X_u|X_s)$ is set to zero, the evolution of X achieves the balance. The equation merely states the fact that the sum of all transitions per unit time into any state X_u must be balanced by the sum of all transitions from X_u into other states X_s. Gain balances loss, in other words, a steady state.

3.2. **Invariance Behaviours.** From the previous argument, we see that the martingale Markov model gives us an (approximating) equation to measure the variation of state transitions in the system. The equation is valid at any time-point and in any state, but the equation provides no clue about $\mathcal{W}(\cdot|\cdot)$, *the transition probability per unit time.* Now an idea is that if it is possible to extract some information about the statistics of $\mathcal{W}(\cdot|\cdot)$, i.e. first and the second moments, then this information should be able to generate a class of sub-models that mimic the behaviour of the original martingale model. We need to find out under what condition the sub-model is equivalent to the original one, in other words, no loss of information on representing $(\Omega, \mathcal{F}, \mathbb{P})$ via the approximating model.

Let $f(\cdot, \cdot)$ be a function satisfying the *maximum principle up to the second order*. It means that for a compact subset $B \in \mathbb{S}$, at time t, the maximum of $f(t, x)$ in $x \in B$ is found on the boundary of B, ∂B. The simplest example of f is a function in the monotone functional class such that for fixed t, $x < x' \in B$ implies $f(t, x) < f(t, x')$ (or $>$), $\nabla_x f(t, x') \geq 0$ (or $\leq$) and $\triangle_x f(t, x') = 0$ on $B \subset \mathbb{R}$. The extremal value of $f(t, \cdot)$ always exists on the boundary of the domain. Here $\triangle_x$ and ∇_x denote the Laplace and gradient operators on x, respectively.

Claim. (Invariance Fairness, IF) If claim MF is true, then for any f satisfying the maximum principle up to the second order, $f(t, X_t)$ will preserve the fairness on a certain measure. The law of $f(t, X_t)$ will also satisfy the maximum principle.

Remark. The claim is another way of specifying Itô's diffusion problem[7]. But to our best knowledge, there is no econom(etr)ics literature concerning on illustrating the problem on the basis of maximum principle. Understanding the connection between this economics claim

[7]Mathematically, this claim intends to squeeze a stochastic problem to a Partial Differential Equation (PDE) problem so that it is possible for economists to construct and solve a specific analytic problem.

and econometrics model is helpful to assess the potentials of modelling. Before doing serious estimations, testing, or predictions, it is better to realize how far the model can reach!

Theorem 3. *For $X_t \in (\mathbb{S}, \mathcal{S})$ and $f(t, \cdot) \in C_b^\infty$, the following are equivalent:*

(i) If claim IF is true, any $f(t, X_t)$ in $C_b^\infty([0,\infty), \mathbb{S})$ has an approximating model that relies on the information contained in the first two moments of the process $f(t, X_t)$.

(ii) The function $f(t, X_t)$ is an Itô diffusion process with drift and diffusion terms, $(a, b) = (a(X_t), b(X_t))$.

Proof. From (ii) to (i), the proof is trivially applying Itô's calculus.

From (i) to (ii), the proof consists of the following four lemmas: 1. to show that the maximum principle on smooth functions is equivalent to the law of Wiener processes, 2. to show the invariance of the law is preserved on the Wiener's path, 3. to set up the approximation on the Wiener's path by showing that the martingale fairness is preserved, 4. to extend the result to the abstract economy $(\Omega, \mathcal{F}, \mathbb{P})$.

The deterministic element x is analogy with X_t in the stochastic process. □

Lemma 4. *If $f(t, x)$ satisfies maximum principle, then $\nabla_x f(t, x)$ is proportional to $\frac{\partial f}{\partial t}(t, x)$.*

If we specify the proportional factor to $-1/2$, then solution of

$$\frac{\partial f}{\partial t}(t, x) = -\frac{1}{2} \triangle_x f(t, x)$$

is the well-known Wiener process. It illustrates a manifest evidence: a state has the invariance property, regardless of any function on it, regardless of any starting value, regardless of any variational and scaling speed, if it is on the *Wiener's path*[8]. In other words, two paths, $\psi(t)$ and $f(\psi(t))$ evolving along time t, should be measurable under the same measure where the maximum principle satisfies.

In order to formalize the concept of Wiener's path, we need to introduce the path space. Suppose that a series of realizations $\{x_{t_i}\}_{t_i \leq t_N}$ corresponds to t via $x_{t_i} = \psi(t_i)$ for $t_i \leq t_N$. Then $\psi : [0, \infty) \to \mathbb{S}$ is a continuous path with the image on the complete separable space $\mathbb{S}$. A path space $\mathfrak{P}(\mathbb{S}) = C([0,\infty), \mathbb{S})$ is a continuous function space of ψs. The σ-algebra $\mathcal{PB}$ is

$$\mathcal{PB}_s := \sigma\left(\psi(t) : t \in [0, s]\right), \quad s \in [0, \infty)$$

generated by $\psi \in \mathfrak{P}(\mathbb{S}) \mapsto \psi(t) \in \mathbb{S}$. The measure $\mathcal{W}$ for $\mathfrak{P}(\mathbb{S})$ is called the Wiener's measure[9]. Note that in previous section $\mathcal{W}$ is used to denote *the transition probability per*

[8] The argument is supported by some weak solutions of the heat equations. We also show how to solve the heat equations by Fourier method. Please see details in the Appendix.

[9] The formula of Wiener's measure, however, is not a rigorous definition. Because Wiener's measure does not have a closed form expression. One should imagine that the limit of the current formula is the law of Wiener process. The limit of the ratio of $\psi(t) - \psi(s)$ and $t - s$ is the path derivative of ψ.

unit time in (3.1). We will see later that the Wiener's measure is exactly the $\mathcal{W}$ in (3.1). Thus we stick to the notation $\mathcal{W}$. The definition of $\mathcal{W}$ is for a sequence $\{\psi(t_i)\}_{t_i \leq t_N} = \{x_{t_i}\}_{t_i \leq t_N}$:

$$\mathcal{W}(\psi : x_1 \in A_{t_1}, \ldots x_t \in A_{t_N}) =$$

$$\int_{A_{t_1}} \cdots \int_{A_{t_N}} \frac{1}{\sqrt{2\pi(t_1 - t_0)}} e^{-\frac{(x_1-x_0)^2}{2(t_1-t_0)}} \cdots \frac{1}{\sqrt{2\pi(t_N - t_{N-1})}} e^{-\frac{((x_N-x_{N-1})^2}{2(t_N-t_{N-1})}} dx_{t_1} \cdots dx_{t_N}.$$

The measure is tight, namely if $t - s < \epsilon$:

$$\lim_{\epsilon \to 0} \sup_{\psi \in \mathfrak{P}(\mathbb{S})} \sup_{0 \leq s \leq t \leq T} \rho(\psi(t), \psi(s)) = 0$$

for any metric $\rho(\cdot, \cdot)$[10].

Lemma 5. *The invariance of Wiener's measure for any function f on $x_t = \psi(t)$, then $\psi(t) - \psi(s)$ is independent and identical for any $s < t$.*

Independent identical increment $\psi(t) - \psi(s)$ together with martingale will give us a "stochastic constant". Recall the path space $\mathfrak{P}(\mathbb{S})$ and its σ-algebra $\mathcal{PB}$. For an independent identical increment $\psi(t) - \psi(s)$ on $\mathfrak{P}(\mathbb{R})$, the Fourier transform is:

$$\mathbb{E}_{\mathcal{W}}\left[e^{i\xi(\psi(t)-\psi(s))} | \mathcal{PB}_s\right] = \int e^{i\xi\varpi} \frac{1}{\sqrt{2\pi(t-s)}} e^{-\varpi^2/2(t-s)} d\varpi = e^{-\frac{|\xi|^2}{2}(t-s)}$$

where $\varpi = \psi(t) - \psi(s)$. What we want is a martingale and a "constant" under $\mathcal{W}$. From the above equation, it easy to see that we can obtain both of them simultaneously if we shift the element $\exp i\xi\psi(t)$ by a Gaussian factor $\exp|\xi|^2 t/2$:

$$\begin{aligned}\mathbb{E}_{\mathcal{W}}\left[e^{i\xi\psi(t)} e^{\frac{1}{2}|\xi|^2 t} | \mathcal{PB}_s\right] =& e^{\frac{1}{2}|\xi|^2 t} \mathbb{E}_{\mathcal{W}}\left[e^{i\xi\psi(t)-\psi(s)+\psi(s)} | \mathcal{PB}_s\right] \\ =& e^{\frac{1}{2}|\xi|^2 t} e^{-\frac{|\xi|^2}{2}(t-s)} \mathbb{E}_{\mathcal{W}}\left[e^{i\xi\psi(s)} | \mathcal{PB}_s\right] \\ =& \mathbb{E}_{\mathcal{W}}\left[e^{i\xi\psi(s)} e^{\frac{1}{2}|\xi|^2 s} | \mathcal{PB}_s\right] = 1.\end{aligned}$$

Let a triplet denote this martingale on the Wiener's path $\mathcal{W}$:

$$\left(\exp\left[i\xi\psi(t) + \frac{1}{2}|\xi|^2 t\right], \mathcal{PB}_t, \mathcal{W}\right) \tag{3.2}$$

which is constant 1 under the expectation w.r.t. $\mathcal{W}$.

This "stochastic constant" will help us to define the approximation error in terms of martingale representation. As in the deterministic case, suppose we define an integral curve of $\psi(\cdot)$ on a smooth vector field a on $\mathbb{R}$, starting at $x \in \mathbb{R}$. Then the path ψ with $\psi(0) = x$ has

[10]Ascoli-Arzela criterion for compact subset.

such a property:

$$f(\psi(t)) - \int_0^t \langle a, \nabla_x f\rangle(\psi(\tau))d\tau$$

is a *constant*[11] for any $f \in C^\infty$. If there is a stochastic analogue, then we can use this *stochastic constant* to set-up our approximating model. The aim is to maintain a stable "error.

Lemma 6. *The triplet* $\left(f(t,\psi) - \int_0^t \left[\nabla_x f + \frac{1}{2}\triangle_x f\right](\tau,\psi)d\tau, \mathcal{PB}_t, \mathcal{W}\right)$ *is a martingale. In addition, if state moves with velocity* $a(X_t)$ *and volatility* $b(X_t)$*, then*

$$\left(f(t,\psi) - \int_0^t \left[a\nabla_x f + \frac{b}{2}\triangle_x f\right](\tau,\psi)d\tau, \mathcal{PB}_t, \mathcal{W}\right)$$

is also a martingale.

Since the martingale with initial condition $\mathcal{W}(\psi(0) = x) = 1$ completely characterizes $\mathcal{W}$, the above result can be extended to any $\mathbb{P}$ by the *Principle of Accompanying Laws* and *Donsker's Invariance Principle* (Theorem 3.1.14 and 3.4.20, [2]) if and only if $\mathbb{P}$ belongs to the family of all tight measures, $\mathcal{M}(\mathfrak{P}(\mathbb{S}))$. In our set-up, $\mathbb{S}$ is a compact metric space so the collection of $\mathbb{P}(\cdot)$ over $\mathcal{S}$ is tight. Principle of Accompanying Laws says if a sequence in complete separable space with tight measure, the law of this sequence will weakly converge. Donsker's Invariance Principle says for independent increment processes, the convergent law is the law of Wiener process.

Lemma 7. *If* $\mathbb{P} \in \mathcal{M}(\mathfrak{P}(\mathbb{S}))$*, then*

$$\left(f(X_t) - f(X_0) - \int_0^t (\mathbb{A}f)dt, \mathcal{PB}_t, \mathbb{P}\right)$$

is a martingale, where $\mathbb{A} := a(\cdot)\nabla_x + \frac{1}{2}b(\cdot)\triangle_x$.

Proof. The IF claim says that a martingale exists for $f(t, X_t)$ on $(\Omega, \mathcal{F}, \mathbb{P})$. The maximum principle restricts the process to be $\mathcal{PB}_t$-adapted, thus $\mathcal{F} \sim \mathcal{PB}$ and the result holds on $(\mathbb{S}, \mathcal{S})$ with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. □

Remark. Some relevant contents can be found in Chapter 3, 4, 7 in [2]. We follow the same motivation as that in Stroock and Varadhan theorem [3], namely formulating the search for the law of a process in terms of a martingale problem. Stroock and Varadhan theorem points out that solving a SDE in the weak sense induces a probability measure which can solve the martingale problem and the converse assertion is also true.

[11] The constant is the initial value $\psi(0) = x$ from the following ODE problem:

$$\frac{\partial f(\psi(t))}{\partial t} = \langle a, \nabla f\rangle(\psi(\tau)).$$

Theorem 3 basically says that if the evolution of state X satisfies the maximal principle up to second moment, the solution of this martingale problem is also the one for diffusion problem. With IF claim, the original martingale model and the approximating martingale model are equivalent in the weak sense such that probability measures share the same law up to the second moment.

This approximating martingale model is a compromise. Being more specific, given the whole transition contents of X, our attention is only restricted on those transitions that will maintain the maximum principle up to the second order. The reason is that only the transitions satisfying *invariance fairness* can be revealed and identified in the standard economic model. For example, suppose X_t is the hidden state with an evolution function $f(X_t)$ and $h(X_t)$ is the choice function for rational consumers. According to the rationality assumption in economics, the choice $h(X_t)$ is observable only if it satisfies the maximum principle. Thus for breaking down the complexity, only the first two orders' law of $f(X_t)$ is taken into account. It does not mean that the unqualified transitions do not exist, conversely, many transitions in the system have high order features such as complex trading strategies in pricing, multiple correlated options, etc.. What we can state here is that those transition features however are too complex to be embedded in a diffusion model[12]. Therefore, those higher order laws of transitions of X will be assigned to the $\mathbb{P}$-null set in $\mathcal{F}$ in the approximating model.

Remark. The invariance property in lemma 5 indicates an important fact: *under $\mathcal{W}$, future increments are independent of the past and have the same distribution as the initial increment.* It is the reason that we are able to construct a specific expression of the optional process on $\mathcal{Y}_t$.

Remark. The generator $\mathbb{A}$ in lemma 7 is said to be local on C_b^∞ if $\mathbb{A}f(x) = 0$ whenever f vanishes in some neighbourhood of x. For a generator with this property, we note that the positive maximum principle implies a local positive maximum principle.

Corollary 8. *If $a(\cdot)$ and $b(\cdot)$ in $\mathbb{A}$ are bounded and continuous, the weak solution of diffusion problem (a, b) is unique. Then*

$$a(X_t) = \int_{-\infty}^{\infty} x\mathcal{W}(X_t|dx), \quad b(X_t) = \int_{-\infty}^{\infty} x^2\mathcal{W}(X_t|dx),$$

where the Wiener's law $\mathcal{W}(\cdot)$ is the transition probability per unit time.

Now it is clear that transition probability per unit time is exactly the Wiener's measure if we add IF claim to MF claim.

[12]One can define a more complicated model to incorporate these effects, but the cost is to use high order stochastic calculus. In fact, later we will see that the diffusion problem already induces an almost infeasible representation for the conditional density. So far, the complexity level of the problems that depart from the diffusion ones is still not quite clear.

Remark. The uniqueness law of $f(X_t)$ is the necessary condition for regularizing the conditional probability π. It releases the fact that the problem of uniqueness of a SDE solution is expressible in terms of the resolvability of a certain parabolic PDE.[13]

If the uniqueness is satisfied, the evolution $f(X_t)$ will maintain the strong Markovian property as X_t in section 3.1. Strong Markovian property is a condition of defining an exponent process e^W with a standard Wiener process W, see [3]. Loosely speaking, one can think that the approximating model only consider the first two items in the Taylor series of e^W.

3.3. **Indifferent Projection.** By the feature of the characteristic function, the martingale (3.2) together with the initial condition captures all the information, first and the second order moment, of $\mathcal{W}$. This implies that $\mathbb{A}(\cdot)$ captures the first two moments information of the process $f(X_t)$ on the economic model $(\Omega, \mathcal{F}, \mathbb{P})$. In other words, the process $f(X_t)$ is a diffusion type process on the Wiener's path.

In fact, IF claim is nothing but pinning the problem onto the *Wiener's space* $L^2(\mathcal{W})$, a L^2 space with Wiener's measure. To see the argument, we need to use the *martingale representation theorem*. The theorem says that any continuous martingale, i.e.

$$M_{f,t} := \left(f(X_t) - f(X_0) - \int_0^t (\mathbb{A}f)dt, \mathcal{F}_t, \mathbb{P} \right)$$

generated by $\mathcal{W}$, can be written as

$$M_f = \mathbb{E}[M_f] + \int_0^T h_s dW_s$$

with a predictable process h_s. Without loss of generality, we consider the case $\mathbb{E}[M_f] = 0$. The functional space of h is

$$L_T^2 := \left\{ h : h \text{ is } \mathcal{F}_t\text{-previsible and } \mathbb{E}\left[\int_0^T \|h_s\|^2 ds \right] < \infty \right\}.$$

The stochastic integral of h is a map $J : L_T^2 \to L^2(\mathcal{F}_T)$ such that

$$J(h) = \int_0^T h_s dW_s.$$

This map is an isometry as the consequence of the Itô isometry theorem. The image of J of the Hilbert space L_T^2 is complete. Therefore, the martingale M_f and the stochastic integral

[13]The duality is: the existence of a solution to the Cauchy problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mathbb{A}u, \qquad \text{in } [0,\infty) \times \mathbb{S} \\ u(0,\cdot) &= f \qquad \text{in } f \in C_b(\mathbb{S}) \end{aligned}$$

implies the uniqueness, at least for one dimensional marginal distributions, of the solution to the martingale problem.

$J(h) \in L^2(\mathcal{F}_t)$ are isometry. What we emphasize here is that the IF claim carries us to a L^2 space where the classic projection techniques are available.

The last thing we have not exploited is the observable process Y. The process Y is used to reflect the law of X, so the topological structure of Y should contain as much information as X. Given any map h in L^2, the isometry property implies that if $Y = h(t, X_t)$, it can maintain all the information in the martingale M_f. However, some information about $\mathbb{P}$-null set in $\mathcal{Y}$, is not contained in the X but affects the outcome of Y. We use the measurement errors to model this information. The following claim is to specify the law of Y.

Claim. (Independent Accompanist, IA) Suppose the observable process Y is contaminated by an additive noise Wiener process W. The noise process W_t is generated by the information set $\mathcal{F}_t$ but is independent of $h(t, X_t)$. The function $h(\cdot)$ satisfies IF claim.

The Wiener process is often modelled independent[14] of X_t. Thus $\mathcal{Y}_t$ is a larger filtration than$\mathcal{F}_t$, e.g.

$$\mathcal{Y}_t = \sigma(X_s, W_s,\ s \in [0, t]) \vee \mathcal{N},$$

since it allows for the measurability of the noise process. Now we can define the process Y with additive noise term:

$$Y_t = \int_0^t h(X_s) ds + W_t, \qquad t \geq 0. \tag{3.3}$$

Note that this specification is to restrict the process Y in $L^2(\mathcal{Y}_t)$ because

$$\mathbb{E}\left[\int h(X_s)^2 ds\right] < \infty, \quad \text{and} \quad W_t \in L^2(\mathcal{Y}_t).$$

Theorem 9. *If claim MF and IF are true, IA claim induces* (3.3) *for observable process* Y_t*. Suppose* $\mathbb{E}\left[\exp\left(\frac{1}{2}\int h(X_s)^2 ds\right)\right] < \infty$*, then the following statements are true:*

(i) under measure $\tilde{\mathbb{P}}$*,*

$$\left.\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \exp\left(-\int_0^t h(X_s) dW_s - \frac{1}{2}\int_0^t h(X_s)^2 ds\right),$$

Y is independent of X and the motions of X under $\tilde{\mathbb{P}}$-law and under $\mathbb{P}$-law are the same.

(ii) For any $\mathcal{F}_t$-measurable random variable $\varphi(X)$,

$$\tilde{\mathbb{E}}\left[\varphi(X)|\mathcal{Y}_t\right] = \tilde{\mathbb{E}}\left[\varphi(X)|\mathcal{Y}\right]$$

where $\mathcal{Y} = \vee_{t\in\mathbb{R}^+}\mathcal{Y}_t$ and $\mathcal{Y}_t = \sigma(Y_s, s \in [0, t]) \vee \mathcal{N}$.

[14]The dependence between W and X is difficult to eliminate in economics and will cause the endogenous problem. But, technically speaking, this issue is often caused by using a too simple function $h(\cdot)$. Since $h(\cdot)$ here can be highly non-linear, i.e. containing all endogenous effects, it is reasonable to ignore this issue here.

This theorem is an important step to derive a specific form for the optional process in (2.1). It allows us to replace the time dependent σ-algebras $\mathcal{Y}_t$ in the conditional expectations $\tilde{\mathbb{E}}[\varphi(X_t)|\mathcal{Y}_t]$ with a time invariant σ-algebra $\mathcal{Y}$. This enables us to use techniques based on Kolmogorov's conditional expectation which would not be applicable if the conditioning set is time dependent, such as $\mathcal{Y}_t$.

We can modify the IA claim to a little bit stronger version which, however, is a common assumption in economics literature:

Claim. (IA') There is a risk-free measure $\tilde{\mathbb{P}}$ such that for any $\mathcal{F}_t$-adapted random variable $\varphi(X_t)$, given all observations (future, present and past), finding the $\mathcal{Y}_t$-optional projection of $\varphi(X_t)$ is equal to compute $\tilde{\mathbb{E}}[\varphi(X_t)|\mathcal{Y}_t]$; that is, future observations have no influence on making an expectation under the measure $\tilde{\mathbb{P}}$.

This claim has a similar role as the MF claim. MF claim is to ensure the existence of a martingale problem for the state process and IA claim does the same job but for the observable process. The aim is to make the state process X, the generator $\mathbb{A}f$ and the observable process Y comparable.

Remark. All the consequences induced by these three claims, e.g. non-arbitrage, maximum principle, and risk-free measures, should not be alien to most economists and should be acceptable for most dynamic models in econom(etr)ics. Thus, the approximating model that characterizes these claims is absolutely a non-trivial model. However, as we stated, some existing economic phenomena as well as statistical features in observable data have lost. MF and IF claims almost squeeze all the complex strategies/behaviours to the $\mathbb{P}$-null set, IA claim ignores non-Gaussian measurement errors. All these ignored events definitely happen in the real world: hedging (complex trading), bounded rational agents (infeasible equilibrium), co-integrated processes (mis-measuring continuous processes), etc. The more data you have, the more complex events it can reflect. On the other hand, the complexity of the model is restricted by imposing these three claims.

Remark. Throughout these sections, one theme is very clear: the model, as well as the approximating one, in some sense, is a "wrong" model. Our attempt is to approximate the domain of the real economic world even though the world is definitely too complicated for the available set-up. This attempt should be treated as a subjective belief such that we believe that our claims capture most essential economic features and the model is on the right track of approximation. Statistics reflect the evidences revealed by real complex economic systems which may be not in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but we believe our model, $(\Omega, \mathcal{F}, \mathbb{P})$, is approximately right. These two facts do not contradict each other.

Remark. If the current situation is that we know the model absolutely does not coincide with the real statistical evidences. Then a natural question is how is our attitude towards statistical techniques? A rejection from the statistical test is not beyond our expectation but the meaning of such a rejection should be interpreted in a right way. A statistical test is a signal to indicate how imperfect the model is. But the target of modelling is not to please the statistical decision. It is true the model is "wrong" and we should realize it! Then a natural question follows: what is the meaning of using a statistical test if one already notices the invalidity of his model?

4. A Probabilistic Estimate

By the claims in section 3, the model is defined as the following pair (X, Y):

Definition. X is a solution of the martingale problem for $(\mathbb{A}; \pi_0)$; in other words, assume that the distribution of X_0 is π_0 and that the process $M_f = \{M_{f,t}, t \geq 0\}$, where

$$M_{f,t} = f(X_t) - f(X_0) - \int_0^t \mathbb{A} f(X_s) ds, \qquad t \geq 0, \tag{4.1}$$

is an $\mathcal{F}_t$-adapted martingale for any $f \in C_b^\infty$ and $(\mathbb{A} f)(\cdot)$ corresponds to $(a(\cdot), b(\cdot))$ of a diffusion process. Y satisfies the evolution equation 3.3, namely

$$Y_t = \int_0^t h(X_s) ds + W_t, \qquad t \geq 0,$$

with null initial condition.

Our concern is a diffusion-type state model and Gaussian noise corrupted observations. The observations are recorded discretely, our idea is to approximate the weak solution of the Kolmogorov forward equation that associates with $\mathbb{A}$ and then use Bayes' rule to incorporate the past observations, a forward and backward analysis.

Both the forward and backward step base on probability distributions. In the forward step, the probabilistic analysis is simplified to a computationally feasible form of the Kolmogorov forward equation. The feasible form is an approximation operating on a simplified representation of the probability distribution, namely mean and covariance. In the backward step, numerical computation method produces optimal fitting for the conditional distribution with the available information, e.g. data and evolution path.

The motivation of this method closely relates to particle filter [ref here] and Kalman filter [ref here] of diffusion processes. Particle filter is to approximate a posteriori distribution by means of the empirical distribution of a system of weighted particles. Kalman filter obtains the mean and covariance of a posteriori distribution by optimal projection to a subspace (or say recursive least square). Our intention is to combine these two approaches. The approach develops from standard Kalman filter but its covariance matrix is computed by the sample covariance of the particles' evolution path instead of the state covariance in Kalman filter. It avoids evolving the covariance matrix of the density function of the state vectors.

In section 4.1, we show that the estimation bases on the so-called *representation theorem*, which is a"limit" form of Bayes' rule. By conditioning on the first and second moment, the representation becomes two separate equations. In section 4.2, we set up an Markov chain approximation by looking forward and then we use past information to interpolate the approximating model and true observations by looking backward.

4.1. **Kushner-Stratonovich-Pardoux Representation.** The filtering problem in principle is to determine the conditional distribution π_t^ω of the $X(\omega)$ at time t given the information accumulated from observing Y in the interval $[0, t]$. Let π_t be a shorthand for π_t^ω. For any bounded continuous function $\varphi \in C_b(\mathbb{S})$, what we want is to compute

$$\pi_t(\varphi) := \mathbb{E}[\varphi(X_t)|\mathcal{Y}_t], \tag{4.2}$$

a conditional expectation of φ. The function φ is also called test function in the generalized function theory. Especially, when $\varphi = x_i$ and $\varphi = x_i x_j$ we have mean and covariance respectively under π_t.

Theorem 1 and 9 imply that a suitable regularization of $\Pi := \{\pi_t, t \geq 0\}$ will make π_t to be an optional (progressively measurable), $\mathcal{Y}_t$-adapted probability measure-valued process for which (4.2) holds almost surely. Thus our current task is to figure out what this process is. We know that for diffusion process the transition density $q(x', x, t)dt := d\mathbb{Q}_t(X'|X)$ satisfies the following type Kolmogorov forward equation:

$$\frac{\partial q(x', x, t)}{\partial t} = \mathbb{A}^* q(x', x, t), \tag{4.3}$$

where $\mathbb{A}^*$ is the adjoint (or dual) operator of $\mathbb{A}$ such that

$$\mathbb{A}^* q = -\sum_i \frac{\partial(a_i(\cdot)q(\cdot)}{\partial x_i} + \frac{1}{2}\sum_{i,j} \frac{\partial^2(b_{ij}(\cdot)q(\cdot))}{\partial x_i \partial x_j}. \tag{4.4}$$

Thus a natural attempt will be connecting the martingale problem in (4.1) with a diffusion type representation. Theorem 3 tells us that when the process is on the Wiener's path, the solution of a martingale problem associated with the second order differential operator is the solution of the diffusion process. Theorem 9 tells us that Y is on the Wiener's path under $\tilde{\mathbb{P}}$.

Proposition 10. *If* $\tilde{\mathbb{E}}[\varphi(X_t)Z_t|\mathcal{Y}_t]$ *is bounded under* $\tilde{\mathbb{P}}$*-law, where*

$$Z_t = \exp\left(-\int_0^t h(X_s)dW_s - \frac{1}{2}\int_0^t h(X_s)^2 ds\right),$$

then for any $\varphi \in C_b(\mathbb{S})$ the process $\rho_t(\varphi) := \tilde{\mathbb{E}}[\varphi(X_t)Z_t|\mathcal{Y}_t]$ follows

$$\rho_t(\varphi) = \pi_0(\varphi) + \int_0^t \rho_s(\mathbb{A}\varphi)ds + \int_0^t \rho_s(\varphi h)dY_s$$

on $\tilde{\mathbb{P}}$ almost surely.

Proof. Note that Z_t is a $\tilde{\mathbb{P}}$-martingale, then

$$Z_t = \exp\left(-\int_0^t h(X_s)dY_s - \frac{1}{2}\int_0^t h(X_s)^2 ds\right)$$

since Y_t is a Wiener process under $\tilde{\mathbb{P}}$. By Girsanov theorem

$$Z_t = 1 + \int_0^t Z_t h(X_s)dY_t.$$

Because $\rho_t(\varphi)$ is bounded, Fubini's theorem and Itô's lemma implies

$$d\tilde{\mathbb{E}}[\varphi(X_t)Z_t|\mathcal{Y}_t] = \tilde{\mathbb{E}}[\mathbb{A}\varphi(X_t)Z_t|\mathcal{Y}_t]dt + \tilde{\mathbb{E}}[\varphi(X_t)h(X_s)Z_t|\mathcal{Y}_t]dY_t. \tag{4.5}$$

Taking the integral, we have the result. □

A new measure is constructed under which Y becomes a Brownian motion and π has a representation in terms of ρ by Bayes' rule such that

$$\pi_t(\varphi) = \frac{\rho_t(\varphi)}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]} = \frac{\rho_t(\varphi)}{\exp\left(\int \pi_s(h)dY_s - \frac{1}{2}\int_0^t [\pi_s(h)]^2 ds\right)}. \tag{4.6}$$

Since $\rho_t(\cdot)$ satisfies a linear evolution equation, we expect this will lead to an evolution equation for π.

Theorem 11. *(Kushner-Stratonovich-Pardoux, KSP) For any $\varphi \in C_b(\mathbb{S})$, proposition 10 induces*

$$\pi_t(\varphi) = \pi_0(\varphi) + \int_0^t \pi_s(\mathbb{A}\varphi)ds + \int_0^t \left(\pi_s(\varphi h) - [\pi_s(h)]^2\right)(dY_s - \pi_s(h)ds). \tag{4.7}$$

Proof. From equation (4.6), we have

$$d\left(\frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]}\right) = \frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]}\left(\int \pi_s(h)dY_s - \frac{1}{2}\int_0^t [\pi_s(h)]^2 ds\right) \tag{4.8}$$

which is equivalent to

$$\pi_t(\varphi) = \rho_t(\varphi) \cdot \frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]}.$$

Note that integration by parts implies

$$\rho_t(\varphi) \cdot \frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]} = \int \frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]} d\rho_t(\varphi) + \int \rho_t(\varphi) d\left(\frac{1}{\tilde{\mathbb{E}}[Z_t|\mathcal{Y}_t]}\right).$$

Substituting equation (4.5) and (4.8), we have the result. □

Remark. In proposition 10 and theorem 11, we did not use the representation of π_t (or unnormalized π_t) directly, but a weak form representation of π_t (or unnormalized π_t). Because the representation of π_t, as in diffusion process case, involves the adjoint operator $\mathbb{A}^*$ as in (4.4), which is not tractable in many cases.

Equation (4.7) is called KSP which is applied to solve non-linear filtering and smoothing problems recently in applied mathematics [10]. One can think KSP representation characterizes an *equilibrium conditional expectation* over any $\varphi \in C_b(\mathbb{S})$. It is a stochastic PDE problem and has a unique solution [15]. Although solving KSP problem can be transferred to solving a parabolic PDE problem, except for the case when the model and the observations are linear and all the disturbances and the initial conditions are normally distributed, finding a closed form expression for these density functions of (4.7) is virtually impossible.

Unlike the state-of-the-art analytic approach, Kalman (linear and non-linear) filter is a rather matured alternative and is familiar to many scholars in control/filtering related fields in economics. The simple scheme of Kalman filter is to update the mean and the covariance of the conditional distribution. In short, it is a sequential or recursive estimation method that is integrated forward in time and, whenever measurements are available, these are used to re-initialize the model before the integration continues. One can see that this scheme is quite similar to the content of KSP equation. In fact, if $h(X_t)$ and $f(X_t)$ at every time t can be linearised as matrices (vectors) $\mathbf{H}_t X_t + \mathbf{h}_t$ and $\mathbf{F}_t X_t + \mathbf{f}_t$ such that

$$X_t = X_0 + \int_0^t (\mathbf{F}_s X_s + \mathbf{f}_s) ds + \int_0^t \sigma_s dV_s, \tag{4.9}$$

$$Y_t = \int_0^t (\mathbf{H}_s X_s + \mathbf{h}_s) ds + W_t, \tag{4.10}$$

then KSP with test functions $\varphi = x_i$ and $\varphi = x_i x_j$ will give us the standard Kalman or called Kalman-Bucy filter.

Proposition 12. *Let $\hat{x}$ be the conditional mean of X such that*

$$\hat{x}_{i,t} = \mathbb{E}[X_{i,t}|\mathcal{Y}_t] = \pi_t(x_{i,t})$$

and R be the conditional covariance such that

$$R_t^{ij} = \mathbb{E}[X_{i,t}X_{j,t}|\mathcal{Y}_t] - \mathbb{E}[X_{i,t}|\mathcal{Y}_t]\mathbb{E}[X_{j,t}|\mathcal{Y}_t] = \pi_t(x_{i,t}x_{j,t}) - \pi_t(x_{i,t})\pi_t(x_{j,t}.)$$

[15] Please refer to chapter 4.8 in [10] for the details about the SPDE problem.

If (4.9) *and* (4.10) *are acceptable localizations for* (4.1)*, then the solution of* $\hat{x}_t$ *satisfies the following SDE*

$$d\hat{x}_t = (\mathbf{F}_t\hat{x}_t + \mathbf{f}_t)dt + R_t\mathbf{H}_t^T(dY_t - (\mathbf{H}_t\hat{x}_t + \mathbf{h}_t)dt), \tag{4.11}$$

and R_t *satisfies the deterministic Riccati equation*

$$\frac{dR_t}{dt} = \sigma_t\sigma_t^T + \mathbf{F}_tR_t + R_t\mathbf{F}_t^T - R_t\mathbf{H}_t^T\mathbf{H}_tR_t. \tag{4.12}$$

Proof. Substituting $\varphi = x_i$ and $\varphi = x_ix_j$ into KSP equation respectively, one will have the result. A detailed proof is given in theorem 4.4.1 [10]. □

Equation (4.11) together with (4.12) is often called the forward *prediction step*. Usually $R_t\mathbf{H}_t^T\mathbf{H}_tR_t$ in (4.12) is expressed as $\mathbf{K}_tC_t^W\mathbf{K}_t$ where $\mathbf{K}_t = R_t\mathbf{H}_t^T[C_t^W]^{-1}$ and C_t^W is the covariance of W_t. The matrix $\mathbf{K}_t$ is called *Kalman gain*. In the backward analysis, we need to use observable information $\mathcal{Y}_t$ to update the system. The *update step* in Kalman filter basically is to minimize the mean square error. By doing this, one sets up a cost function using Gaussian property:

$$p(x) \propto \exp\left(-\frac{1}{2}(x-\mu)^TQ^{-1}(x-\mu)\right),\ p(y|x) \propto \exp\left(-\frac{1}{2}(y-h(x))^TR^{-1}(y-h(x))\right)$$

where $Q = \sigma^T\sigma = b$ and $\mu = a$. The cost function is log-likelihood of $p(x)$ and $p(y|x)$ over t. It is obvious that the optimal cost function relates to recursive least square minimization. We skip the derivation and just list the result:

$$\mathbf{K}_t = R_{t-1}\mathbf{H}_t^T(\mathbf{H}_tR_{t-1}\mathbf{H}_t^T + C_t^W)^{-1},\ \hat{x}_t = \hat{x}_{t-1} + \mathbf{K}_t(y_t - \mathbf{H}_t\hat{x}_{t-1})$$

and $R_t = (I - \mathbf{K}_t\mathbf{H}_t)R_{t-1}$.

There are several obstacles in the Kalman implementation for non-linear system, non-linearization, non-Gaussian, etc. But the most direct thread comes from the conditional covariance R. In both prediction and update step, R is the crucial part of Kalman gain. The fixed point solution of Riccati equation (4.12) is not always available, or even R is solvable, the existence of R^{-1}, which is required in update step, still requires a careful detection.

In this paper, we suggest using sample counterpart covariance instead of solving (4.12) analytically. Applying Monte Carlo method to compute R in Kalman gain is not novel, similar idea can be found in [ref here]. The different thing is that the evolution path of X in our setup comes from KSP representation. The solution of KSP equation is obtained via a stochastic approximation which is given in the next subsection.

Remark. We know that in most case KSP representation certainly induces a non-Gaussian evolution due to the forward integration of non-linear model equations, but we do the localization to truncate those non-linear effects and only use the Gaussian part, first and the second moment, of the distribution just as the spirit of using quasi-likelihood or GMM approach in econometrics.

4.2. **Stochastic Approximation and Interpolation.** We should realize that all approximations to the conditional density of (4.1) are actually approximations to some representation of Bayes' rule, such as the fundamental Bayes' rule formula (4.6). From the perspective of computation, when the representation is involved in computing the approximation to the conditional density, the simplest form of the approximation method is analogous to methods for solving parabolic PDE's by finite differences or finite elements method.

Thus to obtain the representation, we follows the idea of *Euler–Maruyama* approximation, a finite difference numerical method [16] associated with a Markov chain simulator which is used to obtain numerical solution of SDEs. The specific procedure is the following: One constructs an approximating Monte Carlo implementation of the Bayesian update problem on a finite state space, with an approximation parameter τ, and that is "locally consistent" with the diffusion. Then one solves for the cost or optimal cost function for the approximating Markov chain, and finally proves that as $\tau \to 0$, these functions converge to those for the original model.

To stay with a finite state space, we need to adapt the to the discrete X_n, $n \in \mathbb{N}$, for a moment. Let X_n be an $\mathcal{F}_t$-adapted Markov chain. We know that for all t, as $\tau \to 0$, by (3.1) we have

$$\mathbb{Q}_\tau = \mathbb{P}(X_{t+\tau} = j | X_t = i) = \delta_i(j) + q_{ij}(t)\tau + o(\tau)$$

uniformly in t. From equation (4.3), we know that for X_n, the role of (a, b) is taken by the Markov transition matrix $\mathbb{Q}_n = \{q_{ij}(n)\}_{i,j\in\mathcal{I}}$ on finite state space $\mathcal{I}$. The martingale problem (4.1) now becomes

$$M_{t,f} = f(X_t) - f(X_0) - \int_0^t \left(\sum_{j\in\mathcal{I}} q_{ij}(s) f(x_j) | X_s = x_i \right) ds,$$

because $\lim_{\tau\to 0}(\mathbb{Q}_\tau f - f)/\tau = \mathbb{A} f$. The discrete analogue of Y is

$$Y_n = Y_{n-1} + h(t_{n-1}, X_{n-1})\tau + \triangle W_{n-1},$$

and the discrete analogue of Z is

$$Z_n = \exp\left(-\sum_{l=0}^{n-1} h(t_l, X_l)\triangle Y_l - \frac{1}{2}\sum_{l=0}^{n-1} h(t_l, X_l)^2 \triangle t \right),$$

[16] Euler scheme is the most common numerical solver for differential equations.

where $\triangle t$ denotes the time interval for the approximation. Then the representation in (4.6) has the following form:

$$\pi_t^\tau(\varphi) = \frac{\tilde{\mathbb{E}}^\tau[\varphi(X_n)Z_n|\mathcal{F}_t]}{\tilde{\mathbb{E}}^\tau[Z_n|\mathcal{F}_t]},$$

where $\tilde{\mathbb{E}}^\tau$ is taken w.r.t. the simulated law $\tilde{\mathbb{P}}^\tau$. The law of Y_n is Brownian motion under $\tilde{\mathbb{P}}^\tau$. Thus for π_t^τ, only X_n and $h(t_n, X_n)$ matter. We will see later that even $h(\cdot,\cdot)$ is unknown, $h(t_n,\cdot)$ does not matter us because it will be a localization in terms of H_{t_n} while H_{t_n} will be interpolated by some pre-allocated X_n. Therefore, our main concern is the diffusion approximation X_n, a discrete-time finite-state Markov chain whose "local properties" are "consistent" with those of (4.1), namely X_n converges to X_t in the mean square.

Proposition 13. *If the following local consistency conditions hold:*

$$a^\tau(x)\triangle t = a(x)\triangle t + o(\triangle t), \quad b^\tau(x)\triangle t = b(x)\triangle t + o(\triangle t),$$

then

$$\lim_{\tau\to 0}\sup_{t\le T}|\pi_t^\tau(\varphi) - \pi_t(\varphi)| \to 0,$$

for $\varphi \in C_b(\mathbb{S})$.

Proof. In appendix. □

Suppose for each process X_n we generate N sample paths, with equation (4.11), then we have a Monte Carlo estimated moment $\hat{x}_n^\tau$ which means an average of model states evolves in the state space with the mean as the best estimate. Instead of using (4.12), we suggest to use the covariance of the sample paths as the error variance R_n^τ. Since the sample paths are generated independently, we know $\hat{x}_n^\tau$ and R_n^τ are unbiased estimates of $\hat{x}_t$ and R_t when $n = t$. The information contained in first and the second moment of π_t becomes available now.

Before showing the full scheme, we need to give an illustration about how to interpolate the unknown $h(\cdot)$. The idea is to extrapolate (predict) a value of h and then fit the function of h at observation points by one step ahead extrapolation and one step behind interpolation. This idea can be found in other works, e.g. [9]. As an illustration, let us consider the simple case of three grid points e, f, g and two observations, 1 and 2, see figure 4.1. Monte Carlo approximation generates X_e, X_f and X_g where g is an extrapolation point, then we have interpolated observations, $h_0(X_e)$, $h_0(X_f)$ and $h_0(X_g)$, where $h_0(\cdot)$ could be an initial linear function or some other basis functions. With $h_0(X_e)$, $h_0(X_f)$ and $h_0(X_g)$ and observations Y_1 and Y_2, we can tune the coefficient values of the interpolated function $h_0(\cdot)$ and then use it for next step analysis.

The full scheme of this estimate is listed below:

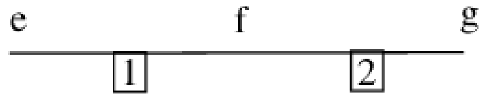

FIGURE 4.1. Illustration about extrapolation and interpolation

(1) Creation of the N initial states: If initial condition X_0 and R_0 are known, we generate N number of normal distributed random variables. The k-th draw has the following form

$$X_0^\tau(k) = X_0 + R_0^{\frac{1}{2}}\varepsilon_0$$

where ε_0 is the standard normal variable.

(2) On n step, create forward states and compute sample mean and covariance: On discrete time interval $[n-1, n]$, for each $k \leq N$, we use Markov chain $\mathbb{Q}$ to generate the forward state of X_{k-1}^τ

$$X_n^\tau(k) = X_{n-1}^\tau(k) + a(X_{n-1}^\tau(k)) + b(X_{n-1}^\tau(k))(\varepsilon_n - \varepsilon_{n-1}).$$

The forward sample mean is

$$\overline{X_n^\tau} = \frac{1}{N}\sum_{k=1}^{N} X_n^\tau(k).$$

The forward variance is

$$\overline{R_n^\tau} = \frac{1}{N}\sum_{k=1}^{N}(X_n^\tau(k) - \overline{X_n^\tau})(X_n^\tau(k) - \overline{X_n^\tau}).$$

(3) On step n, create backward estimated states and compute sample mean and covariance: Use simulated X_n^τ to extrapolate $\mathbf{H}_n$. By (4.11) and (4.12), we have backward states

$$\hat{x}_n(k) = X_n^\tau(k) + \mathbf{K}_n[Y_n(k) - \mathbf{H}_n X_n^\tau(k)]$$

where

$$\mathbf{K} = \overline{R_n^\tau}\mathbf{H}_n^T[\mathbf{H}_n\overline{R_n^\tau}\mathbf{H}_n^T + C_n^W]^{-1}$$

is the sample counterpart Kalman gain. The mean and covariance of backward states are

$$\hat{x}_n^\tau = \frac{1}{N}\sum_{k=1}^{N}\hat{x}_n(k)$$

and

$$R_n^\tau = (I - \mathbf{K}_n\mathbf{H}_n)\overline{R_n^\tau}(I - \mathbf{K}_n\mathbf{H}_n)^T + \mathbf{K}_n C_n^W \mathbf{K}_n^T.$$

(4) Set the mean and covariance of backward states as the initial state on step $n+1$.

The estimate method tries to break those seemingly infeasible problems, solving KSP for the whole conditional density and updating covariances based on the second moment evolution of KSP representation, into several tractable problems. Two potential concerns for the method are worth to mention. Firstly, though the method uses the full non-linear dynamics to propagate the forecast error statistics, it mimics the traditional Kalman filter in the analysis step and uses only the Gaussian part of the prior distribution. We know that least-squares estimation is very inefficient for highly non-Gaussian processes, whose distributions are not well characterized by means and variances. Secondly, the updated sample paths preserve only the first two moments of the "posterior". Consequently, the initial condition for the further integration of the Kolmogorov forward equation does not coincide with the "posterior" one, an inconsistent scheme. We leave these questions to the future research.

5. Inference For A Simulated Complex System

> Why linear systems are so important. The answer is simple: because we can solve them! - Richard Feynman
>
> The fluttering of a butterfly's wing in Rio de Janeiro, amplified by atmospheric currents, could cause a tornado in Texas two weeks later. - Edward Lorenz

The most famous complex system is the deterministic 3-states Lorenz system (the butterfly effect):

$$\begin{aligned}\frac{dx_1}{dt} &= \theta_1(x_2 - x_1),\\ \frac{dx_2}{dt} &= \theta_2 x_1 - x_2 - x_1 x_3,\\ \frac{dx_3}{dt} &= x_1 x_2 - \theta_3 x_3.\end{aligned}$$

The original derivation of these equations are for fluid flow of the atmosphere: a two-dimensional fluid cell is warmed from below and cooled from above and the resulting motion is modeled by this system. Later on, the system is used to model some of the unpredictable behaviour which normally associate with chaotic phenomena in the economics, finance, biology, earth science, physics and even philosophy of human brain. One can also found some empirical evidences from lower dimensional cases of Lorenz system in economics, for example, logistic map in cobweb models, see [ref here].

One of the most important differences between chaotic processes and truly stochastic processes is that the future behaviour of a chaotic system can be predicted in the short term, while stochastic processes are characterized in terms of their statistics. Since in practice none of economic data has truly predictable features, the deterministic Lorenz model should be embedded into a stochastic model. The procedure is simple, one simply add stochastic disturbances into each of these functions:

$$\begin{aligned}
dX_1 &= \theta_1(X_2 - X_1)dt + dW_1, \\
dX_2 &= (\theta_2 X_1 - X_2 - X_1X_3)dt + dW_2, \\
dX_3 &= (X_1X_2 - \theta_3 X_3)dt + dW_3.
\end{aligned}$$

The system is called stochastic Lorenz system or stochastic Lorenz attractor. From figure 5.2, one can see the system is still approximately attached to a regular manifold as in the deterministic case, but the movement path becomes completely irregular.

We model the non-linear feature of $h(\cdot)$ by using a random matrix $\mathbf{H}$ at each time step. The reason is that at each time t, if there is no drift, then linearisation of $h(x_t)$, $\mathbf{H}_t x_t + Rem$, gives us Gaussian property of $\mathbf{H}_t x_t$ and truncates the non-Gaussian effect by ignoring the reminder term Rem; however, the propagation of this effect still disturbs the future linearisation and therefore attaches the non-Gaussian features to the system. The simplest way to model this non-linear feature is to use a time varying random coefficient matrix $\mathbf{H}_t(\omega)$. The process Y is assumed to be discretely observed, every five movements of X generates one Y:

$$\mathbf{Y}_t = \mathbf{H}_t(\omega)\overline{\mathbf{X}}_t + \varepsilon_t$$

where $\mathbf{Y} = (Y_1, Y_2, Y_3)^T$, $\mathbf{X} = (X_1, X_2, X_3)^T$, $\mathbf{H}_t(\omega) = \mathbf{H}_{t-1}\mathbf{X}_{t-1} - \mathbf{Y}_{t-1} + \beta\varepsilon$, and ε is standard normal. The notation $\overline{\mathbf{X}}_t$ means an average of five moments of $\mathbf{X}_t$ in the time interval $[t-1, t]$. In the experiment, we try two different values of β, please see figure 5.6 and 5.7.

When solving the system, it is not very difficult to perform the iterations and then take the average over various realizations. However, the accuracy of Euler-Maruyama method is not very good. In particular it is very hard to improve the accuracy by an order of magnitude, therefore we use Dormand–Prince method[17]. Although the evolution of Y depends on this highly non-linear mechanism, in prediction step, we assume Y to be Gaussian. Thus the computational speed is significantly reduced but the estimated model still maintains an acceptable level of root mean square errors, please see figure 5.4.

Appendix:Proof of Theorem and related Lemma

Before giving the proof of theorem 1, we will introduce three lemmas. The result of theorem 1 will be the consequence of these lemmas. Lemma 14 constructs a countable vector space $\mathcal{U}$ on $\mathbb{S}$. Lemma 15 defines a non-negative process that corresponds to the element in $\mathcal{U}$, Lemma f16 extends $\mathcal{U}$ to the space of continuous bounded functions, $C_b(\mathbb{S})$, checks whether

[17]A build-in matlab solver, *ode45,* uses this method.

the definition of the process is still valid, and find a representation of π_t Finally, theorem 1 use extend the results in $C_b(\mathbb{S})$ to $B(\mathbb{S})$.

Lemma 14. *For any continuous bounded choice function $\varphi(x)$ where $x \in \mathbb{S}$ and $\mathbb{S}$ is a compact set, there is a representation of $\varphi(\cdot)$ consisting of a class of orthogonal choice functions $\{\varphi_i\}_{i=1}^{\infty}$ such that $\varphi_i \subset C_b(\mathbb{S})$.*

Proof. For $C_b(\mathbb{S})$, compact $\mathbb{S}$ induces that $C_b(\mathbb{S})$ is dense and that a linear span exists. Let $\{\varphi_i\}_{i=1}^{\infty}$ be the set of basis function in the linear span and thus any φ_i is bounded continuous. Let $\mathcal{U}$ be a countable vector space generated by finite linear combinations of $\{\varphi_1, \dots, \varphi_n\}$ with rational coefficients such that

$$\mathcal{U} := \left\{ \varphi = \sum_{i=1}^{n} \alpha_i \varphi_i,\ a_i \text{ is rational for all } i \right\},$$

where $\varphi_1 = 1$. These φ_is are linearly independent for any $i \in \mathbb{Z}^+$. □

Lemma 15. *There is a $\mathcal{Y}_t$-measurable process $\Lambda_\omega^t(\varphi)$ corresponding to the choice function of stochastic states $\varphi(X_t)$.*

Proof. For $\omega \in \Omega$ and any fixed t, $X_t(\omega)$ is a function $X_t(\omega) : \omega \mapsto \mathbb{S}$. Lemma 14 gives a representation of φ on the base $\mathcal{U} := \{\varphi = \sum_{i=1}^{n} \alpha_i \varphi_i\}$. Thus for any $\varphi_i \in \mathcal{U}$ and fixed t, $\sum_{i=1}^{n} \alpha_i \varphi_i(X_t)$ is a $\mathcal{F}_t$-measurable representation of φ. [18]. Equation (2.1) implies that a $\mathcal{Y}_t$-adapted representation g_n^t exists for $\varphi_n(X_t)$. Thus a sequence $\{g_i^t\}_{i=1}^{n}$ is corresponding to $\{\varphi_i(X_t)\}_{i=1}^{n}$. Thus if $\sum_{i=1}^{n} \alpha_i \varphi_i(X_t)$ represent $\varphi(X_t)$, then a $\mathcal{Y}_t$-adapted process g^t, corresponding to φ, is linearly and uniquely represented by $\sum_{i=1}^{N} \alpha_i g_i^t$. We can define the linear functional

$$\Lambda_\omega^t(\varphi) = g^t(\omega), \quad \forall t.$$

□

Because the conditional distribution is a non-negative process, we need to construct a non-negative analogue of Λ_ω^t.

Lemma 16. *There is a non-negative $\mathcal{Y}_t$-measurable process $\tilde{\Lambda}_t^\omega(\varphi)$ corresponding to any continuous and bounded choice function $\varphi \in C_b(\mathbb{S})$. Furthermore, the non-negative $\tilde{\Lambda}_t^\omega(\varphi)$ has a representation w.r.t. the non-negative distribution process π_t^ω such that $\tilde{\Lambda}_t^\omega(\varphi) = \langle \pi_t^\omega, \varphi \rangle$.*

Proof. Define a countable sub-base for positive choice function:

$$\mathcal{U}^+ := \left\{ \varphi = \sum_{i=1}^{n} \alpha_i \varphi_i,\ \varphi \geq 0 \right\}.$$

[18] This statement skips one intermediate assumption which requires X_t to be progressive measurable (see [2] Remark 7.1.1 Lemma 7.1.2).

For $\varphi \in \mathcal{U}^+$ and fixed t, we define the null set for φ such that

$$\mathcal{N}(\varphi) := \left\{\omega \in \Omega : \ \Lambda^t_\omega(\varphi) < 0\right\},$$

where $\Lambda^t_\omega(\varphi)$ is the $\mathcal{Y}_t$-measurable process from lemma 15. If $\varphi(X_t) \geq 0$ almost surely, then by equation (2.1) the optional process would be non-negative on $\mathcal{U}^+$ and hence $\mathcal{N}(\varphi)$ is a $\mathbb{P}$-null set for $\mathcal{U}^+$. To extend this construction to $\mathcal{U}$, we can define a new process $\bar{\Lambda}^\omega_t$:

$$\bar{\Lambda}^\omega_t(\varphi) := \begin{cases} \Lambda^\omega_t(\varphi) & \omega \notin \mathcal{N}(\varphi), \\ 0 & \omega \in \mathcal{N}(\varphi). \end{cases}$$

Next, we need to check whether $\bar{\Lambda}^\omega_t$ is bounded. It is obvious that $\bar{\Lambda}^\omega_t(1) = 1$. Since $\varphi \in \mathcal{U}$, the uniform norm has such a property that $|\varphi| \leq \|\varphi\|_\infty 1$. Then $\|\varphi\|_\infty 1 \pm \varphi \geq 0$, from step 2, we know

$$\begin{aligned} \bar{\Lambda}^\omega_t(\|\varphi\|_\infty 1 \pm \varphi) \geq 0 \\ \|\varphi\|_\infty \pm \bar{\Lambda}^\omega_t(\varphi) \geq 0 \end{aligned}$$

where the second inequality comes from the linearity of $\bar{\Lambda}^\omega_t$ and $\bar{\Lambda}^\omega_t(1) = 1$. It implies

$$\sup_t \|\bar{\Lambda}^\omega_t(\varphi)\|_\infty < \|\varphi\|_\infty,$$

so $\bar{\Lambda}^\omega_t(\varphi)$ is bounded for $\varphi \in C_b(\mathbb{S})$.

Let any $\varphi \in C_b(\mathbb{S})$. Since $\mathcal{U}$ is dense in $C_b(\mathbb{S})$, there exists a sequence $\varphi_k \in \mathcal{U}$ such that $\varphi_k \to \varphi$. We can define

$$\tilde{\Lambda}^\omega_t(\varphi) := \begin{cases} \bar{\Lambda}^\omega_t(\varphi) & \varphi \in \mathcal{U}, \\ \lim_k \Lambda^\omega_t(\varphi_k) & \varphi \in C_b(\mathbb{S}) \setminus \mathcal{U}. \end{cases}$$

over the all $C_b(\mathbb{S})$. For boundedness, we only need to check the case $\varphi \in C_b(\mathbb{S}) \setminus \mathcal{U}$. Note that for any two sequence φ_k and φ_j, if $\varphi_k \to \varphi$ and $\varphi_j \to \varphi'$, we will have

$$\sup \|\tilde{\Lambda}^\omega_t(\varphi_k) - \tilde{\Lambda}^\omega_t(\varphi_j)\|_\infty \leq \|\varphi_k - \varphi\|_\infty + \|\varphi - \varphi'\|_\infty + \|\varphi' - \varphi_j\|_\infty$$

by the boundedness result in $\mathcal{U}$ and the triangle inequality. So $\tilde{\Lambda}^\omega_t(\varphi)$ is bounded.

We also need to ensure that the optional process of $\tilde{\Lambda}^\omega_t(\varphi)$ is well-defined on $C_b(\mathbb{S})$. For φ_k in $\mathcal{U}$, we have $\mathcal{Y}_t$-adapted process $\tilde{\Lambda}^\omega_t(\varphi_k)$ for $\varphi_k(X_t)$, then

$$\begin{aligned} \mathbb{E}\left[\tilde{\Lambda}^\omega_T(\varphi)\mathbb{I}_{T<\infty}\right] &= \lim_{k\to\infty} \mathbb{E}\left[\tilde{\Lambda}^\omega_T(\varphi_k)\mathbb{I}_{T<\infty}\right], \\ &= \lim_{k\to\infty} \mathbb{E}\left[\varphi_k(X_T)\mathbb{I}_{T<\infty}\right], \\ &= \mathbb{E}\left[\varphi(X_T)\mathbb{I}_{T<\infty}\right]. \end{aligned}$$

The last equation comes from the dominated convergence theorem for bounded sequence.

Since $\mathbb{S}$ is compact, Riesz representation theorem shows the existence of π_t^ω,

$$\tilde{\Lambda}_T^\omega(\varphi) = \int_{\mathbb{S}} \varphi(x)\pi_t^\omega(dx) = \langle \pi_t^\omega, \varphi \rangle = \pi_t^\omega \varphi, \qquad \text{for } \forall t$$

for any bounded and well-defined inner product. □

Proof of Theorem 1.

Proof. We need to extend the definition of $\pi_t^\omega \varphi$ in lemma 16 to incorporate $\varphi \in B(\mathbb{S})$. Let $\bar{B}(\mathbb{S})$ is a subset of $B(\mathbb{S})$ such that $\pi_t^\omega \varphi$ is $\mathcal{Y}_t$-adapted optional process of $\varphi(X_t)$ on $\bar{B}(\mathbb{S})$. It is obvious that $C_b(\mathbb{S}) \subset \bar{B}(\mathbb{S})$. Note that the Borel σ-algebra generated by $B(\mathbb{S})$ is $\mathcal{B}(\mathbb{S})$. By the completeness of $C_b(\mathbb{S})$, we can construct a sequence of subset $\{\bar{B}_i(\mathbb{S})\}_i^\infty$ such that

$$\bar{B}_1(\mathbb{S}) \subset \bar{B}_2(\mathbb{S}) \subset \cdots .$$

Compactness of $\mathbb{S}$ implies that $\mathcal{B}(\mathbb{S})$ is closed under finite intersection. From the construction in step 1, we know the constant function is included in every $\bar{B}_i(\mathbb{S})$. Monotone class theorem implies $\cup_i \bar{B}_i(\mathbb{S}) \supseteq \mathcal{B}(\mathbb{S})$, since any monotone non-negative increasing sequence $\{\bar{B}_i(\mathbb{S})\}_i^\infty$, with indicator function of every set in $\mathbb{S}$, contains the σ-algebra $\mathcal{B}(\mathbb{S})$ which is closed under finite intersection. Thus $\bar{B}(\mathbb{S})$ contains every bounded $\mathcal{S}$-measurable function of $\mathbb{S}$. While $\bar{B}(\mathbb{S})$ is a subset of $B(\mathbb{S})$, we conclude $\bar{B}(\mathbb{S}) = B(\mathbb{S})$. □

Proof of Lemma 4.

Proof. The maximum principle definition is simply the first and the second derivative conditions in the calculus. If a function $f : \mathbb{S} \to \mathbb{R}$ attains its maximum at point $x \in \mathbb{S}$, then

$$\nabla_x f(t,x) = 0 \qquad \text{and} \qquad \triangle_x f(t,x) \leq 0$$

Furthermore, since f is a time-dependent function such that $f : [0,T) \times \mathbb{S} \to \mathbb{R}$ at a certain time interval $[0,t]$, f attains its maximum on x when time is t, then

$$\frac{\partial f}{\partial t}(t,x) \geq 0.$$

Together with $\nabla_x f(t,x) = 0$ and $\triangle_x f(t,x) \leq 0$, the inequality $\partial_t f \geq 0$ express the uncertainty of the future such that $\partial_t f(x,\cdot)$ could either strictly increase along t or obtain its optimal at t. A naïve attempt of formalizing the idea of preserving the maximum principle on $([0,T) \times \mathbb{S})$ is to connect these two inequality functions by an equality[19]:

$$\frac{\partial f}{\partial t}(t,x) = -\frac{1}{2}\triangle_x f(t,x), \tag{5.1}$$

which is the heat equation. Without loss of generality, we consider the standard case with the diffusion factor $-1/2$ but the result holds for any vector factor $-b/2$. □

[19]In physics, the equality is set by Fourier's heat conduction law.

Proof of Lemma 5.

Proof. Identical: Note that a function f over ψ will not change the expression except that $\psi(t)$ is replaced by $f(\psi(t))$. By Lemma 3.4.3 and Theorem 3.4.16 (Kolmogorov's Criterion) [2], Stroock shows that for a subset μ of all tight measures $\mathcal{M}(\mathfrak{P}(\mathbb{S}))$ and $\psi \in \mathfrak{P}(\mathbb{R})$:

$$\sup_{\mu \in \mathcal{M}(\mathfrak{P}(\mathbb{S}))} \mathbb{E}_\mu \left[|\psi(t) - \psi(s)|^r \right] \leq C_T |t-s|^{1+\alpha},$$

where $C_T < \infty$ is a constant, $\alpha > 0$ and $r \geq 1$. Then we have

$$\lim_{t \to s} \sup_{\psi \in \mathfrak{P}(\mathbb{S})} \frac{(\psi(t) - \psi(s))^2}{(t-s)} = \lim_{t \to s} \sup_{\psi \in \mathfrak{P}(\mathbb{S})} \left(\frac{\psi(t) - \psi(s)}{t-s} \right)^2 (t-s) \to 0.$$

It means the increments are controlled by the length of time interval. When the interval is extremely small, all the increments are treated the same. So the smooth function f does not matter the law of $\mathcal{W}$.

Independent: For ψ, $\varpi \in \mathfrak{P}(\mathbb{R})$, let $\varpi(t) = \psi(t+s) - \psi(s)$, by the formula of Wiener's measure, both $\psi(s)$ and $\varpi(t)$ associate with $\mathcal{W}$ on the time path $[0, s]$ and $[0, t]$ respectively. Clearly, they are independent. □

Proof of Lemma 6.

Proof. We define the Fourier transform of f by $\mathbb{F}f(\xi) = \int_{-\infty}^{\infty} f(x) e^{i\xi x} dx$, and the inverse Fourier transform is $\mathbb{F}^{-1} f(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx$.[20].

As in the deterministic case, the ideal representation of $f(t, \psi(t))$ on $\mathcal{W}$ is the path integral:

$$\int_0^t \left[\nabla_x f + \frac{1}{2} \triangle_x f \right] (\tau, \psi(\tau)) d\tau.$$

We need to check whether the approximation error is a "constant" in the stochastic sense. Note that

$$f(t, x) = (2\pi)^{-1} \int e^{i(\xi t + \xi x)} (\mathbb{F}^{-1} f) d\xi d\eta.$$

By the property $\mathbb{F}^{-1}(\frac{\partial}{\partial x})(\cdot) = i\xi \mathbb{F}^{-1}(\cdot)$, there is

$$\mathbb{F}^{-1} \left(\nabla_x f + \frac{1}{2} \triangle_x f \right) = \left(i\xi - \frac{1}{2} |\xi|^2 \right) (\mathbb{F}^{-1} f).$$

[20]More discussion about Fourier method is given in the Appendix.

The approximating error is

$$f(t,\psi(t)) - \int_0^t \left[\nabla_x f + \frac{1}{2}\triangle f\right](\tau,\psi(\tau))d\tau$$
$$= (2\pi)^{-1}\int\int \underbrace{\left[e^{i(\xi t+\xi\psi(t))} - \int_0^t e^{i(\xi\tau+\xi\psi(\tau))}(i\xi - \frac{1}{2}|\xi|^2)d\tau\right]}_{M_\xi(t)}(\mathbb{F}^{-1}f)d\xi d\eta$$

The Fourier term $\mathbb{F}^{-1}f$ is bounded and irrelevant to $\mathcal{W}$, if $M_\xi(t)$ is martingale in $\mathcal{W}$, then the error will be a stochastic constant. Rewrite $M_\xi(t)$ as:

$$M_{\xi,\eta}(t) = e^{i\xi t}e^{i\xi x} - \int_0^t e^{i\xi\psi(\tau)}e^{i\xi\tau}d(i\xi - \frac{1}{2}|\xi|^2)\tau.$$

The second term can be written as

$$\int_0^t e^{i\xi\psi(\tau)+\frac{1}{2}|\xi|^2\tau}d(e^{i\xi\tau}\cdot e^{-\frac{1}{2}|\xi|^2\tau})$$

and the first term can be written as $e^{i\xi t-\frac{1}{2}|\xi|^2 t}e^{i\xi\psi(t)+\frac{1}{2}|\xi|^2 t}$. Fubini's Lemma together with (3.2) implies that

$$\mathbb{E}_{\mathcal{W}}\left[M_\xi(t)|\mathcal{PB}_s\right] = 1\cdot\mathbb{E}_{\mathcal{W}}\left[e^{i\xi t-\frac{1}{2}|\xi|^2 t} - \int_0^t d(e^{i\xi\tau-\frac{1}{2}|\xi|^2\tau})d\tau|\mathcal{PB}_s\right] = 1.$$

Thus $\left(f(t,\psi) - \int_0^t\left[\nabla_x f + \frac{1}{2}\triangle_x f\right](\tau,\psi)d\tau, \mathcal{PB}_t, \mathcal{W}\right)$ is a martingale.

If the state moves with velocity $a(X_t)$, the path derivative becomes $a(\cdot)\nabla f$. Moreover, the Lapalace operator $\triangle$ in the heat equation may associate with a volatility coefficient $b(\cdot)$. Then the approximating model is:

$$\int_0^t\left[a(X_s)\nabla_x f + \frac{1}{2}b(X_s)\triangle_x f\right]ds$$

which is the integral of the Feller's generator $\mathbb{A}$ on f:

$$\mathbb{A} := a(\cdot)\nabla_x + \frac{1}{2}b(\cdot)\triangle_x.$$

The generator is a dual representation of a diffusion process (a,b) such that

$$dX_t = a(X_t)dt + \sigma(X_t)dV_t$$

where $b(X_t) = \sigma(X_t)^T\sigma(X_t)$ and V_t is a Wiener process. $\square$

Proof of Theorem 9.

Proof. (i) The bounded condition

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int h(X_s)^2 ds\right)\right] < \infty,$$

is called Novikov's condition. By this condition, Girsanov's theorem implies Z_t

$$\left.\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}\right|_{\mathcal{F}_t} = Z_t := \exp\left(-\int_0^t h(X_s)dW_s - \frac{1}{2}\int_0^t h(X_s)^2 ds\right)$$

is an $\mathcal{F}_t$-adapted martingale and Martingale representation theorem implies

$$W_t + \left\langle \int_0^t h(X_s)dW_s, W_t \right\rangle_t = W_t + \int_0^t h(X_s)ds = Y_t,$$

where $\langle \cdot, \cdot \rangle_t$ is the quadratic variation such that $\langle W_t, W_t \rangle_t = t$. Thus for $d\tilde{\mathbb{P}} = Z_t d\mathbb{P}$, Y_t is a Brownian motion with respect to $\tilde{\mathbb{P}}$:

$$\begin{aligned} &\mathbb{E}e^{\left(W_t + \int_0^t h(X_s)ds\right)} e^{\left(-\int_0^t h(X_s)dW_s - \frac{1}{2}\int_0^t h(X_s)^2 ds\right)} \\ &= \mathbb{E}e^{\left\{\int_0^t (1+h(X_s))dW_s - \int_0^t (2h(X_s)+h^2(X_s))ds\right\}} \\ &= \mathbb{E}e^{t^2/2} \cdot e^{\left\{\int_0^t (1+h(X_s))dW_s - \int_0^t (1+h(X_s))^2 ds\right\}} = e^{t^2/2}. \end{aligned}$$

The last line is the result of (3.2).

The law of the pair process (X, Y) can be written as

$$(X_t, Y_t) = (X_t, W_t) + \left(0, \int_0^t h(X_s)ds\right),$$

thus on an arbitrary time interval $[0, t]$, under $\tilde{\mathbb{P}}$-law, the law of (X_t, W_t) is absolutely continuous with respect to the law of the pair process (X_t, Y_t). For any bounded measurable function φ defined on the product path space of (X, Y), we have

$$\tilde{\mathbb{E}}\left[\varphi(X_t, Y_t)\right] = \mathbb{E}\left[\varphi(X_t, Y_t)Z_t\right] = \mathbb{E}\left[\varphi(X_t, W_t)\right].$$

Therefore, X and Y are independent under $\tilde{\mathbb{P}}$ since X and W are independent.

(ii) Under the probability measure $\tilde{\mathbb{P}}$, the law of the process Y is completely specified as a $\mathcal{F}_t$-adapted Wiener process with independent increments of Y. Hence, the σ-algebra is $\mathcal{Y}_t^\dagger = \sigma(Y_{t+u} - Y_t)$ for any $u \geq 0$. Note that $\mathcal{Y}_t$ and $\mathcal{Y}_t^\dagger$ are independent. By the condition expectation property:

$$\tilde{\mathbb{E}}\left[\varphi(X_t)|\mathcal{Y}_t\right] = \tilde{\mathbb{E}}\left[\varphi(X_t)|\sigma(\mathcal{Y}_t, \mathcal{Y}_t^\dagger)\right].$$

Since $\mathcal{Y}_t^\dagger$ includes all the increment information after time t,

$$\sigma(\mathcal{Y}_t, \mathcal{Y}_t^\dagger) = \mathcal{Y}_t \vee \mathcal{Y}_{(t'-t)\in\mathbb{R}} = \mathcal{Y},$$

and $\mathcal{Y}$ is a time invariant σ-algebra. $\square$

Itô integral. Suppose a process $\phi_n(t,\omega) : [0,T] \times \Omega \to \mathbb{R}$ has such a partition

$$0 = t_0 < t_1 < \cdots < t_{n+1} = T$$

that $\phi_n(s,\omega) = h_j(\omega)$ when $t_j < s < t_{j+1}$ where $h_j(\omega)$ is a random variable. $h_j(\omega)$ is required to be measurable with respect to the filtration $\mathcal{F}_t$ such that X_t is martingale w.r.t. $\mathcal{F}_t$. Itô integral is:

$$\int_0^T \phi_n(t,\omega)dX_t = \sum_{j=0}^{n} h_j(\omega)(X_{t+1} - X_t).$$

A more general stochastic process can be defined as limit of integrals of simple processes such that

$$\int_0^T f(t,\omega)dX = \lim_{n\to\infty} \int_0^T \phi_n(t,\omega)dX_t.$$

Four solution forms of heat equation. *Four solution forms of the heat equation: Suppose $f(t,x)$ is a solution of* (5.1) *and suppose $g(t)$ is any differentiable function that*

$$\int_A g(y)f(t,x-y)dy < \infty,$$

then (i)

$$v(t,x) = (g \star f)(t,x) = \int_A g(y)f(t,x-y)dy,$$

(ii) $v(t,x) = f(t,x-y)$, (iii) $v(t,x) = f_t(t,x)$ or $v(t,x) = f_x(t,x)$, (iv) $v(t,x) = f(a^2t,ax)$ for $a \in \mathbb{R}$, are all solutions of (5.1)*. The symbol $\star$ means convolution.*

It is straightforward to show the fact. (i)

$$\frac{\partial v}{\partial t}(t,x) - \triangle_x v(t,x) = \int_A \left[\frac{\partial f}{\partial t}(t,x-y) - \triangle_x f(t,x-y)\right] g(y)dy = 0,$$

(ii) Change the variable $z = x - y$ then $\partial_t f(t,z) = \partial_t f(t,x)$ and$\partial_z^2 f(t,z) = \partial_x^2 f(t,x)$. (iii)

$$\frac{\partial v}{\partial t}(t,x) - \triangle_x v(t,x) = \frac{\partial}{\partial t}\left[\frac{\partial f}{\partial t}(t,x-y) - \triangle_x f(t,x-y)\right] = \frac{\partial}{\partial t}0 = 0,$$

(iv) $\partial_t v(t,x) = a^2\partial_t f(a^2t,ax)$, $\partial_x v(t,x) = a\partial_x f(a^2t,ax)$,

$$\frac{\partial^2}{\partial x^2}v(t,x) = a\frac{\partial}{\partial x}\partial_x f(a^2t,ax) = a^2\frac{\partial^2}{\partial x^2}f(a^2t,ax)$$

then

$$a^2\left[\frac{\partial f}{\partial t}(t,x-y) - \frac{\partial^2}{\partial x^2}f(a^2t,ax)\right] = 0.$$

Fourier analysis and Fourier solution of Heat equation. A "weak" solution of heat equations is given here. "Weak" here means that solution should, presumably, have a distribution type representation instead of classic ones, i.e. x or $x^2 - t$. To do that, we need an initial condition, $f(0, x)$.

Proof. A periodic function $f \in C^\infty$ of period T has a Fourier series[21]:

$$f(x) = \sum_{-\infty}^{\infty} a(\xi_k) e^{i\xi_k x}(\xi_k - \xi_{k-1}), \qquad a(\xi_k) = \frac{1}{2\pi}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(x) e^{-i\xi_k x} dx$$

where $\xi_k = 2\pi k/T$. If $f(x)$ is not periodic, let $T \to \infty$ the expression becomes

$$f(x) = \int_{-\infty}^{\infty} a(\xi) e^{i\xi x} dx, \qquad a(\xi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx.$$

Two series f and a can represent each other. We define the Fourier transform of f by $\mathbb{F}f(\xi) = \int_{-\infty}^{\infty} f(x) e^{i\xi x} dx$, and the inverse Fourier transform is $\mathbb{F}^{-1} f(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i\xi x} dx$. Note that $\mathbb{F}^{-1}\mathbb{F} f = \mathbb{F}\mathbb{F}^{-1} f = f$. Now apply Fourier transform to (5.1) over the variable x:

$$\frac{\partial}{\partial t}\mathbb{F}f(t, \xi) = -\frac{1}{2}\xi^2 \mathbb{F}f(t, \xi)$$

by the fact that the Fourier transform of $\partial_x f(t, x)$ is $-i\xi \mathbb{F}f(t, \xi)$. Solving this ODE we have

$$\mathbb{F}f(t, \xi) = c(\xi) e^{-\xi^2 t/2}$$

where $c(\xi) = \mathbb{F}f(0, \xi)$ by the Fourier transform of the initial condition. Therefore:

$$\begin{aligned} f(x, t) =& \mathbb{F}^{-1}\left[\mathbb{F}f(0, \xi) e^{-\xi^2 t/2}\right] \\ =& \frac{1}{\sqrt{2\pi t}} \int e^{-(x-y)^2/2t} f(0, y) dy. \end{aligned}$$

If the initial condition is $f(0, x) = \delta_x$, then the solution is the Gaussian distribution. □

Proof of Proposition 13. The convergence of

$$\pi_t^\tau(\varphi) - \pi_t(\varphi) = \frac{\tilde{\mathbb{E}}^\tau[\varphi(X_n) Z_n | \mathcal{F}_t]}{\tilde{\mathbb{E}}^\tau[Z_n | \mathcal{F}_t]} - \frac{\tilde{\mathbb{E}}[\varphi(X_t) Z_t | \mathcal{F}_t]}{\tilde{\mathbb{E}}[Z_t | \mathcal{F}_t]}$$

essentially depends on the convergence of Z_n. The quadratic term in

$$Z_n = \exp\left(\sum_{l=0}^{n-1} h(t_l, X_l) \triangle Y_l - \frac{1}{2}\sum_{l=0}^{n-1} h(t_l, X_l)^2 \triangle t\right)$$

is easy to deal with when $\triangle t \to 0$. So we only consider the stochastic summation term.

[21]Weierstrass' second approximation theorem: every continuous function of period 2π is uniformly approximable by trigonometric polynomials.

Because $\exp\left(\sum_{l=0}^{n-1} h(t_l, X_l)\triangle Y_l\right) \to \exp\int h(X_l)dY_l$ when $\triangle t \to 0$. We will study the limit case only. Let ς_1 and ς_2 be two independent bounded processes that are independent of two standard Wiener process W_1 and W_2. Let

$$D = \mathbb{E}\sup_{t\leq T}\left|\exp\left[\int_0^t \varsigma_1(s)dW_1(s)\right] - \exp\left[\int_0^t \varsigma_2(s)\right]\right|.$$

We will use the inequality

$$\left|e^A - e^B\right| \leq |A-B|(e^A+e^B).$$

By Schwarz's inequality and the above inequality

$$\begin{aligned}D^2 \leq &\mathbb{E}\sup_{t\leq T}\left|\left[\int_0^t (\varsigma_1(s)-\varsigma_2(s))dW_1(s)\right]\right|^2 \times \\ &\left[\mathbb{E}\sup_{t\leq T}\left|\exp\left[\int_0^t (\varsigma_1(s))dW_1(s)\right] + \exp\left[\int_0^t (\varsigma_2(s))dW_2(s)\right]\right|^2\right].\end{aligned}$$

The first part on the RHS of the inequality is bounded by martingale inequality[22]

$$\mathbb{E}\sup_{t\leq T}\left|\left[\int_0^t (\varsigma_1(s)-\varsigma_2(s))dW_1(s)\right]\right|^2 \leq 4\mathbb{E}\left[\int_0^t (\varsigma_1(s)-\varsigma_2(s))ds\right],$$

and the second part is bounded by martingale inequality and Itô isometry:

$$\begin{aligned}&\mathbb{E}\sup_{t\leq T}\left|\exp\left[\int_0^t (\varsigma_1(s))dW_1(s)\right] + \exp\left[\int_0^t (\varsigma_2(s))dW_2(s)\right]\right|^2 \\ &\leq 4\mathbb{E}\left|\exp\left[\int_0^t (\varsigma_1(s))dW_1(s)\right] + \exp\left[\int_0^t (\varsigma_2(s))dW_2(s)\right]\right|^2\end{aligned}$$

and for $i = 1, 2$

$$\mathbb{E}\exp\left[\int_0^t \varsigma_i(s)dW_i(s)\right]^2 \leq \mathbb{E}\exp\left[\int_0^t |\varsigma_i(s)|^2 ds\right].$$

Hence we have

$$D^2 \leq C_1\mathbb{E}\int_0^T \left|\varsigma_1(s)-\varsigma_2(s)\right|^2 ds < \infty.$$

Then $\sup_{t\leq T}\mathbb{E}|Z_n - Z_t|$ is bounded, especially,

$$\mathbb{E}|Z_n - Z_t|^2 \to 0$$

when $\tau \to 0$, because by local consistency condition the Markov chain $\mathbb{Q}$ converges in the mean square to $\mathbb{A}$.

The local consistency condition implies

$$\sup_{t\leq T}|\mathbb{E}^\tau\varphi(X_n)Z_n - \mathbb{E}\varphi(X_t)Z_t| \to 0$$

[22] For any real value (sub)martingale M, there is $\mathbb{E}\sup_{t\leq T}M^2 \leq 4\mathbb{E}M^2$.

as $\tau \to 0$. Then

$$\lim_{\tau\to 0}\sup_{t\leq T}|\pi_t^\tau(\varphi)-\pi_t(\varphi)| \leq \lim_{\tau\to 0}\sup_{t\leq T}\mathbb{E}\left|\varphi(X_n)-\varphi(X_t)\right| Z_n$$
$$+C_2\lim_{\tau\to 0}\sup_{t\leq T}\mathbb{E}|Z_n-Z_t| \to 0.$$

References

[1] Freddy Delbaen, Walter Schachermayer, The Mathematics of Arbitrage, 2006, Springer.

[2] Daniel W. Stroock, Probability Theory: an Analytic View, 2000, Cambridge University Press.

[3] Daniel W. Stroock, S. R. S. Varadhan, Diffusion Processes with Continuous Coefficients, I, 1969, 22(3), 345-400, Communications on Pure and Applied Mathematics.

[4] Andrey Kolmogorov, Foundations of the Theory of Probability (2nd Ed), 1956, Chelsea.

[5] L. C. G. Rogers, David Williams, Diffusions, Markov Processes and Martingales: Volume 2, 2000, Cambridge University Press.

[6] Nico van Kampen, Stochastic Processes in Physics and Chemistry (3rd Ed), 2007, North Holland.

[7] Dean Corbae, Maxwell B. Stinchcombe, Juraj Zeman, An Introduction to Mathematical Analysis for Economic Theory and Econometrics, 2009, Princeton University Press.

[8] Jing Lei and Peter Bickel, 2009, Ensemble Filtering for High Dimensional Non-linear State Space Models, Working Paper.

[9] Durham, Garland B., and A. Ronald Gallant, 2002, Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes, The Journal of Business and Economic Statistics 20, 297-316.

[10] Alain Bensoussan, Stochastic Control of Partially Observable Systems, 2004, Cambridge University Press.

[11] Olav Kallenberg, Foundations of Modern Probability, 2002, Springer.

[12] Lars Hansen, Y. Ait-Sahalia and J. Scheinkman, 2009, Operator Methods for Continuous-Time Markov Processes, Handbook of Financial Econometrics, Vol 1, Chapter 1, Elsevier Press.

[13] William Brock and Cars Hommes, 1997, A Rational Route to Randomness, Econometrica, 65, 5, 1059-1095.

[14] George W. Evans and Garey Ramey, 1992, Expectation Calculation and Macroeconomic Dynamics, American Economic Review, 82, 1, 207-224.

[15] Marcet, and Sargent, 1989, Convergence of Least Squares Learning in Environments with Hidden State Variables and Private Information, Journalof Political Economy, 97, 1306-1322.

[16] Hansent and Sargent, 2007, Robustness, Princeton University Press.

[17] Steven D. Levitt, John A. List and David H. Reiley, 2010, What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments, Econometrica, 78, 4, 1413-1434.

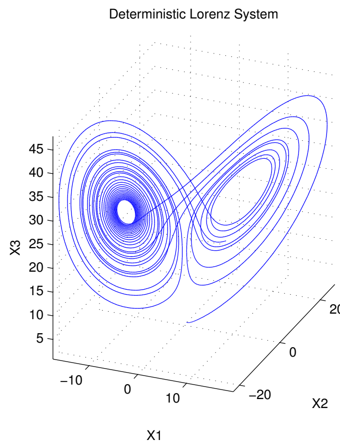

FIGURE 5.1. Deterministic Lorenz System: $\theta_1 = 28$, $\theta_2 = 10$, $\theta_3 = 8/3$.

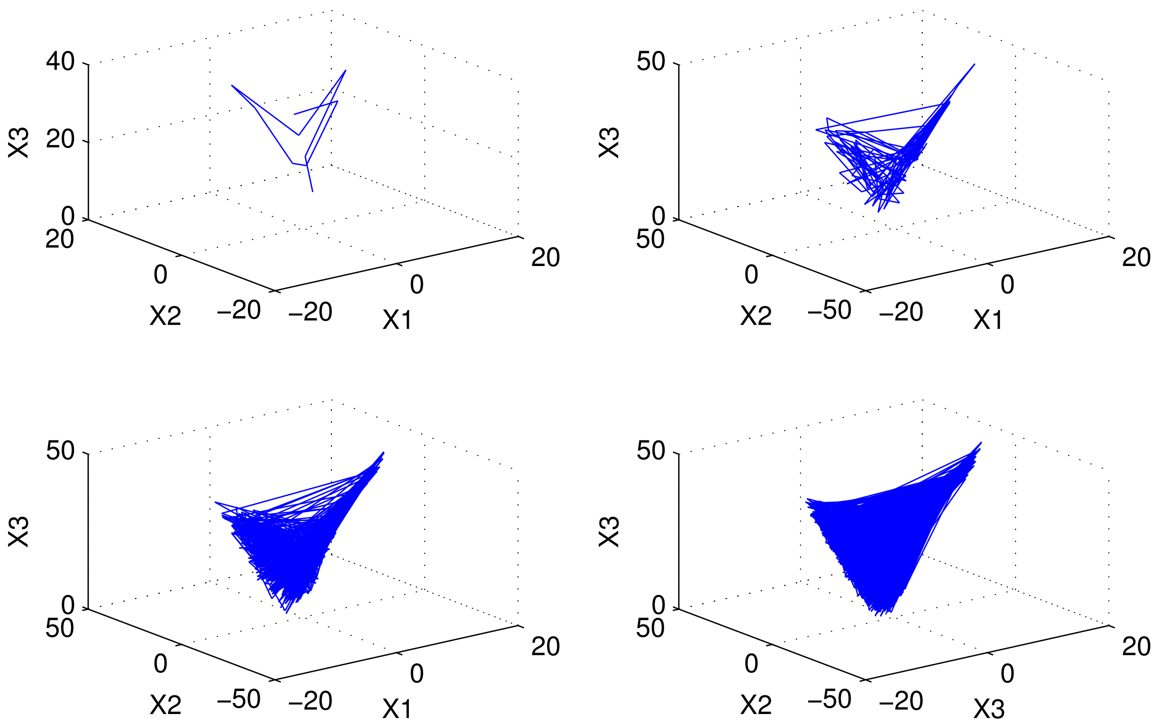

FIGURE 5.2. Stochastic Lorenz System: Top left to bottom right, $T = 10$, $T = 100$, $T = 1000$, $T = 10000$.

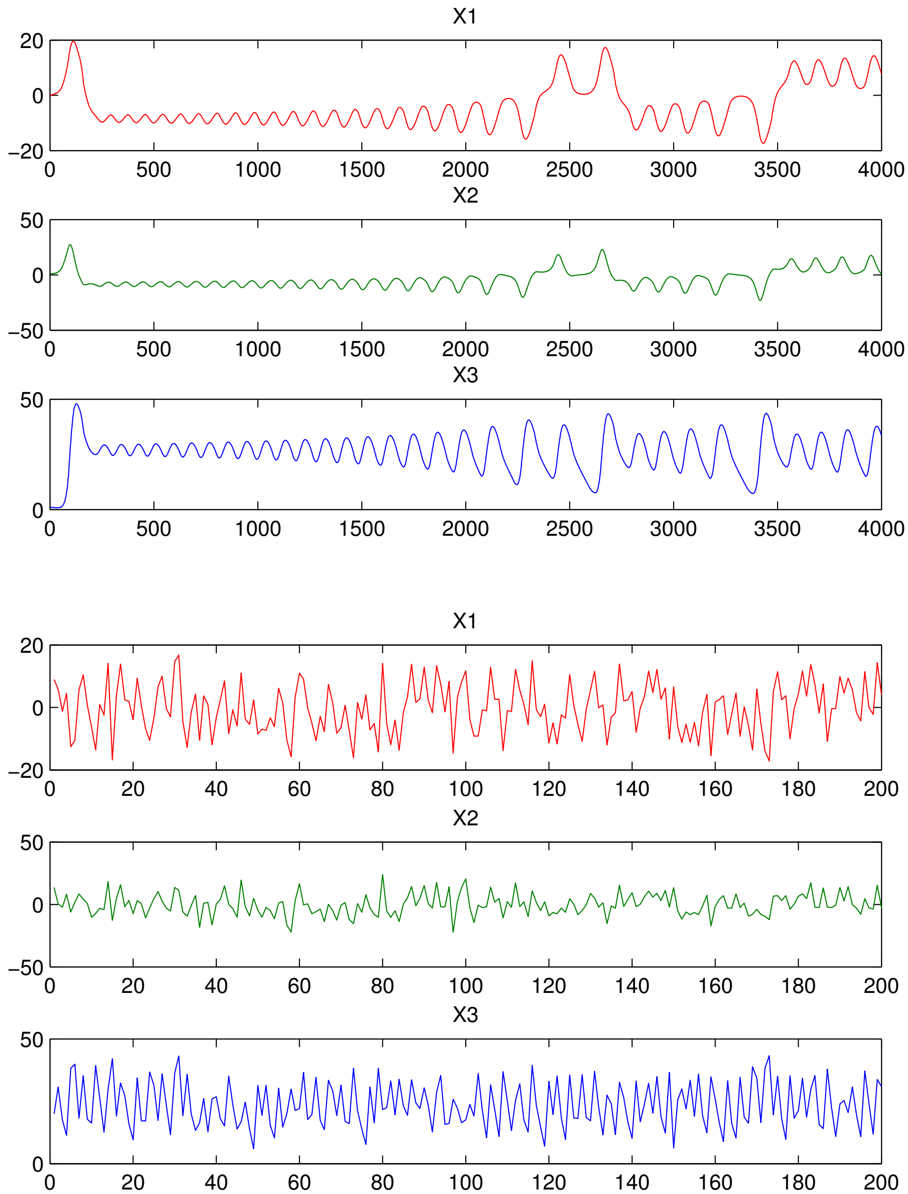

FIGURE 5.3. Left: First 4000 evolutions for x_1, x_2, x_3. Right: First 200 evolutions for X_1, X_2 and X_3.

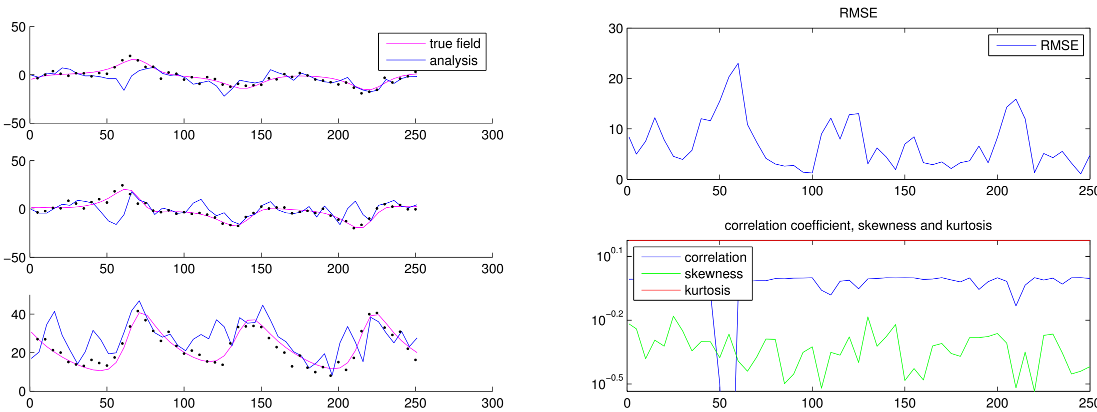

FIGURE 5.4. Estimation with interpolation

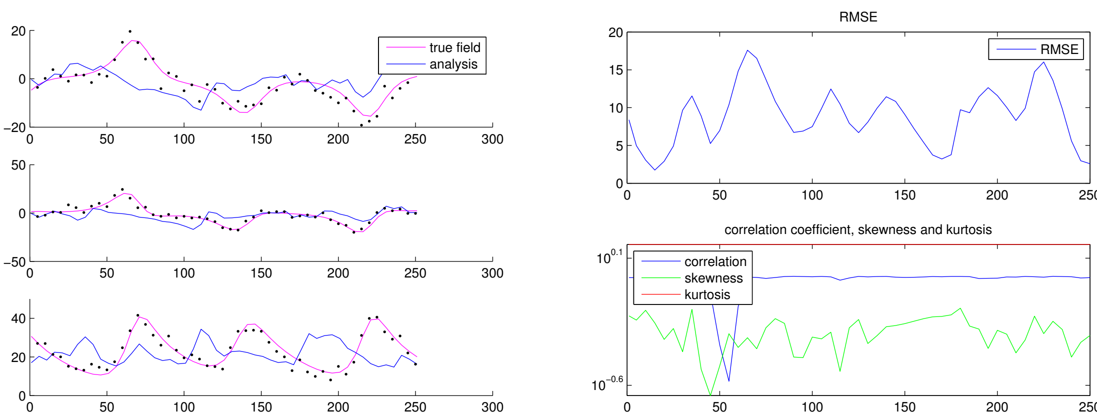

FIGURE 5.5. Estimation without interpolation

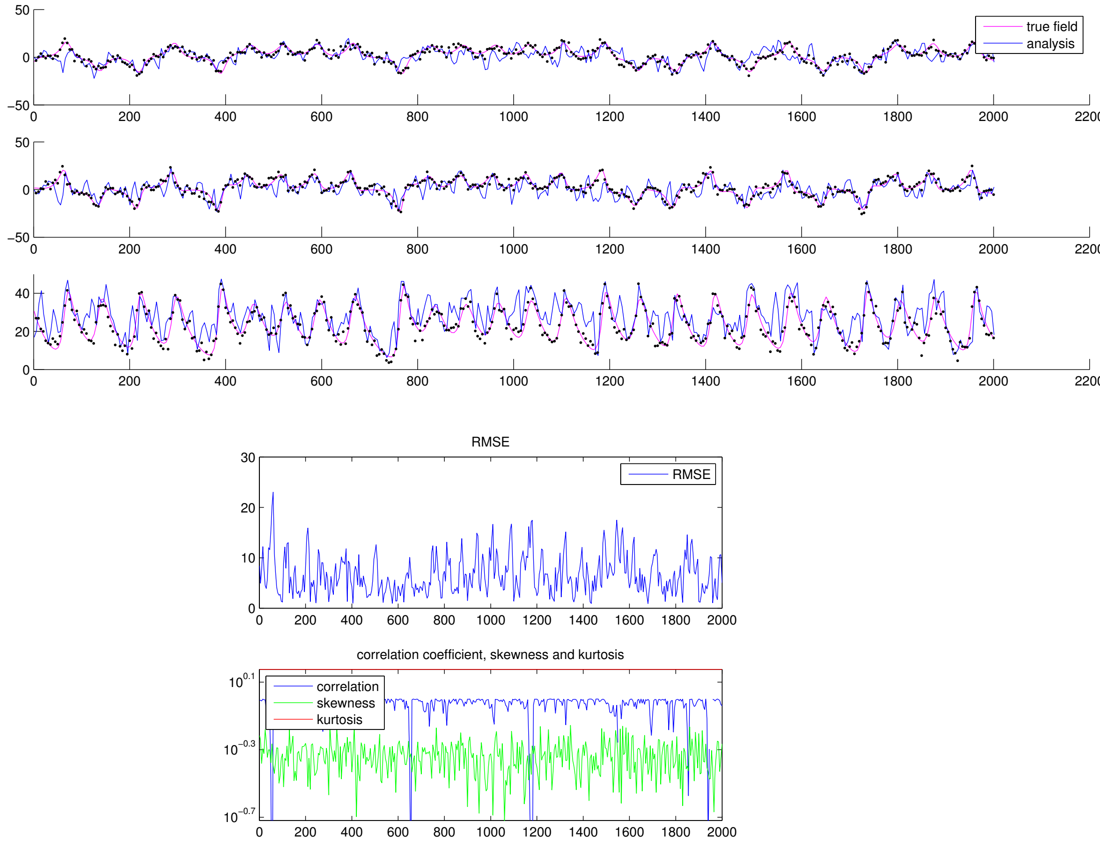

FIGURE 5.6. Estimation with interpolation, $\beta = 9$.

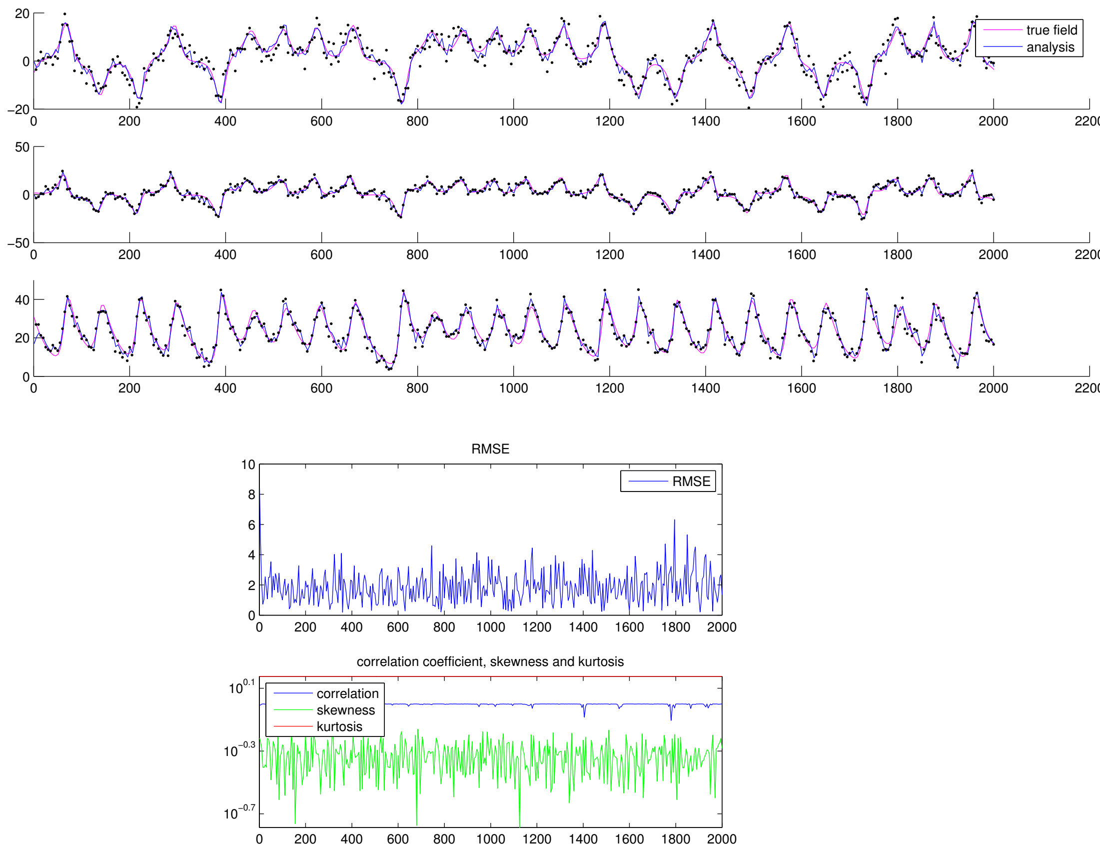

FIGURE 5.7. Estimation with interpolation, $\beta = 1$.