Parameter Estimation for Rough Differential Equations

Anastasia Papavasiliou*
and
Christophe Ladroue†

December 10, 2008

Abstract

We construct an estimator based on "signature matching" for differential equations driven by rough paths and we prove its consistency and asymptotic normality. Note that the the Moment Matching estimator is a special case of this estimator.

Key words: Rough paths, Diffusions, Generalized Moment Matching, Parameter Estimation.

1 Motivation

Consider the following system:

$$dY_t^{1,\epsilon} = \frac{1}{\epsilon} f(Y_t^{1,\epsilon}) Y_t^{2,\epsilon} dt, \quad Y_0^{1,\epsilon} = y \tag{1}$$

$$dY_t^{2,\epsilon} = -\frac{1}{\epsilon} Y_t^{2,\epsilon} dt + \frac{1}{\sqrt{\epsilon}} dW_t, \quad Y_0^{2,\epsilon} = y' \tag{2}$$

and the process

$$d\bar{Y}_t = f(\bar{Y}_t) dW_t, \quad \bar{Y}_0 = y. \tag{3}$$

defined on a probability space $(\Omega, (\mathcal{F}_t)_t, \mathbb{P})$ where W is Brownian motion and $t \in [0, T]$. Under certain assumptions on f, it is possible to show that $Y^{1,\epsilon}$ converges to $\bar{Y}$ in distribution as $\epsilon \to 0$ (see [2]). We call $\bar{Y}$ the homogenization limit of $Y^{1,\epsilon}$.

Suppose that we know that f is a smooth function ($f \in \mathcal{C}(\mathbb{R})$) and we are asked to estimate it. What we have available is a "black box", which is usually some complicated computer program or, possibly, a lab experiment, whose output is $\{Y_t^{1,\epsilon}\}_{t\in[0,T]}$.

*Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.
†Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK.

We input the initial conditions and we get realizations of $Y^{1,\epsilon}$ corresponding to the unknown f and the given initial conditions. However, because of the multiscale structure of the model, simulating long paths is computationally very demanding. Such situations often arise in molecular dynamics or atmospheric sciences. The problem we just posed is a simplified version of problems arising in these fields. Recently, a number of algorithms have been developed to deal with such problems. They come under the title of "equation-free" (see [1] and references within). The main idea is to use short simulations of the microscale model (1) in order to locally estimate the macroscale model (3). Applying this idea to the example described above, we get the following algorithm:

(i) *Initialization:* For $n = 0$ (first loop), set $Y_0^{1,\epsilon} = y_0$ for some arbitrary y_0. Otherwise, for $n > 0$, set $Y_0^{1,\epsilon} = y_n$.

(ii) *Estimation:* Run several independent copies of the system initialized according to step (i). Fit the data $\{Y_t^{1,\epsilon}(\omega_i)\}_{t\in[0,T]}$ where for $i = 1, \ldots, N$ these are independent realizations of $Y^{1,\epsilon}$, to the following model:

$$Y_t^{1,\epsilon} = \frac{1}{\epsilon}\left(\sum_{k=0}^{p} \frac{a_k}{k!}(Y_t^{1,\epsilon} - Y_0^{1,\epsilon})^k\right) Y_t^{2,\epsilon} dt$$

derived from (1) by replacing f by a polynomial approximation of some degree p. If $\hat{a}_k$ are the estimates of the parameters of the polynomial, approximate $f^{(k)}(Y_0^{1,\epsilon})$ by $\hat{a}_k$, where $f^{(k)}(Y_0^{1,\epsilon})$ is the k^{th} derivative of f at $Y_0^{1,\epsilon}$.

(iii) *Taking a step:* If $p = 0$, set $y_{n+1} = y_n + \delta$ for some $\delta > 0$. Otherwise set $y_{n+1} = \hat{f}_{n,p}(\delta)$, where $\hat{f}_{n,p}$ is the local polynomial approximation around y_n that corresponds to the estimated first p derivatives of f at y_n. For example, for $p = 1$, $y_{n+1} = y_n + \hat{a}_1\delta$. After taking the step, go to (i).

We repeat this algorithm several times, until we have enough information about f. We do a polynomial fit, using all estimated derivatives.

In the rest of the paper, we focus on how to do the estimation required in step (ii) above. The exact mathematical question that we will try to answer is described in the following section.

2 Setting

2.1 Some basic results from the theory of Rough Paths

In this section, we review some of the basic results from the theory of rough paths. For more details, see [3] and references within.

Consider the following differential equation

$$dY_t = f(Y_t) \cdot dX_t,\ Y_0 = y_0. \tag{4}$$

We think of X and Y as paths on a Euclidean space: $X : I \to \mathbb{R}^n$ and $Y : I \to \mathbb{R}^m$ for $I := [0,T]$, so $X_t \in \mathbb{R}^n$ and $Y_t \in \mathbb{R}^m$ for each $t \in I$. Also, $f : \mathbb{R}^m \to \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$, where $\mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ is the space of linear functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ which is isomorphic to the space of $n \times m$ matrices. We will assume that $f(y)$ is a polynomial in y. We will also assume that X is a path of finite p-variation, meaning that

$$\sup_{\mathcal{D}\subset[0,T]} \left(\sum_{\ell} \|X_{t_\ell} - X_{t_{\ell-1}}\|^p \right)^{\frac{1}{p}} < \infty,$$

where $\mathcal{D} = \{t_\ell\}_\ell$ goes through all possible partitions of $[0,T]$ and $\|\cdot\|$ is the Euclidean norm. Note that we will later define finite p-variation for multiplicative functionals.

What is the meaning of (4) when $p \geq 2$? In order to answer this question, we first need to give some definitions:

Definition 2.1. *Let $\Delta_T := \{(s,t);\ 0 \leq s \leq t \leq T\}$. Let $p \geq 1$ be a real number. We denote by $T^{(k)}(\mathbb{R}^n)$ the k^{th} truncated tensor algebra*

$$T^{(k)}(\mathbb{R}^n) := \mathbb{R} \oplus \mathbb{R}^n \oplus \mathbb{R}^{n\otimes 2} \oplus \cdots \oplus \mathbb{R}^{n\otimes k}.$$

(1) Let $\mathbf{X} : \Delta_T \to T^{(k)}(\mathbb{R}^n)$ be a continuous map. For each $(s,t) \in \Delta_T$, denote by $\mathbf{X}_{s,t}$ the image of (s,t) through $\mathbf{X}$ and write

$$\mathbf{X}_{s,t} = \left(\mathbf{X}^0_{s,t}, \mathbf{X}^1_{s,t}, \ldots, \mathbf{X}^k_{s,t}\right) \in T^{(k)}(\mathbb{R}^n), \text{ where } \mathbf{X}^j_{s,t} = \left\{\mathbf{X}^{(i_1,\ldots,i_j)}_{s,t}\right\}_{i_1,\ldots,i_j=1}^{n}.$$

The function $\mathbf{X}$ is called a **multiplicative functional** *of degree k in $\mathbb{R}^n$ if $\mathbf{X}^0_{s,t} = 1$ for all $(s,t) \in \Delta_T$ and*

$$\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t} = \mathbf{X}_{s,t} \quad \forall s,u,t \text{ satisfying } 0 \leq s \leq u \leq t \leq T,$$

i.e. for every $(i_1, \ldots, i_\ell) \in \{1, \ldots, n\}^\ell$ and $\ell = 1, \ldots, k$:

$$(\mathbf{X}_{s,u} \otimes \mathbf{X}_{u,t})^{(i_1,\ldots,i_\ell)} = \sum_{j=0}^{\ell} \mathbf{X}^{(i_1,\ldots,i_j)}_{s,u} \mathbf{X}^{(i_{j+1},\ldots,i_\ell)}_{u,t}$$

(2) A **p-rough path $\mathbf{X}$** *in $\mathbb{R}^n$ is a multiplicative functional of degree $\lfloor p \rfloor$ in $\mathbb{R}^n$ that has finite p-variation, i.e. $\forall i = 1, \ldots, \lfloor p \rfloor$ and $(s,t) \in \Delta_T$, it satisfies*

$$\|\mathbf{X}^i_{s,t}\| \leq \frac{(M(t-s))^{\frac{i}{p}}}{\beta\left(\frac{i}{p}\right)!},$$

where $\|\cdot\|$ is the Euclidean norm in the appropriate dimension and β a real number depending only on p and M is a fixed constant. The space of p-rough paths in $\mathbb{R}^n$ is denoted by $\Omega_p(\mathbb{R}^n)$.

(3) A **geometric p-rough path** *is a p-rough path that can be expressed as a limit of 1-rough paths in the p-variation distance d_p, defined as follows: for any $\mathbf{X}, \mathbf{Y}$ continuous functions from Δ_T to $T^{(\lfloor p \rfloor)}(\mathbb{R}^n)$,*

$$d_p(\mathbf{X}, \mathbf{Y}) = \max_{1 \leq i \leq \lfloor p \rfloor} \sup_{\mathcal{D} \subset [0,T]} \left(\sum_{\ell} \| \mathbf{X}^i_{t_{\ell-1}, t_\ell} - \mathbf{Y}^i_{t_{\ell-1}, t_\ell} \|^{\frac{p}{i}} \right)^{\frac{i}{p}},$$

where $\mathcal{D} = \{t_\ell\}_\ell$ goes through all possible partitions of $[0,T]$. The space of geometric p-rough paths in $\mathbb{R}^n$ is denoted by $G\Omega_p(\mathbb{R}^n)$.

One of the main results of the theory of rough paths is the following, called the "extension theorem":

Theorem 2.2 (Theorem 3.7, [3])**.** *Let $p \geq 1$ be a real number and $k \geq 1$ be an integer. Let $\mathbf{X} : \Delta_T \to T^{(k)}(\mathbb{R}^n)$ be a multiplicative functional with finite p-variation. Assume that $k \geq \lfloor p \rfloor$. Then there exists a unique extension of $\mathbf{X}$ to a multiplicative functional $\hat{\mathbf{X}} : \Delta_T \to T^{(k+1)}(\mathbb{R}^n)$.*

Let $X : [0,T] \to \mathbb{R}^n$ be an n-dimensional path of finite p-variation for $n > 1$. One way of constructing a p-rough path is by considering the set of all iterated integrals of degree up to $\lfloor p \rfloor$. If $X_t = \left(X_t^{(1)}, \ldots, X_t^{(n)} \right)$, we define $\mathbf{X} : \Delta_T \to T^{(\lfloor p \rfloor)}$ as follows:

$$\mathbf{X}^0 \equiv 1 \in \mathbb{R} \text{ and } \mathbf{X}^k_{s,t} = \left\{ \int \cdots \int_{s < u_1 < \cdots < u_k < t} dX^{(i_1)}_{u_1} \ldots dX^{(i_k)}_{u_k} \right\}_{(i_1, \ldots, i_k) \in \{1, \ldots, n\}^k} \in \mathbb{R}^{n \otimes k}$$

When $p \in [1,2)$, the iterated integrals are well defined. However, when $p \geq 2$ there will be more than one way of defining them. For example, if X in an n-dimensional Brownian motion, the second iterated integrals would be different, depending on whether we use the Itô, the Stratonovich or some other rule. What the extension theorem says is that if the path has finite p-variation and we define the first $\lfloor p \rfloor$ iterated integrals, the rest will be uniquely defined. So, if the path is of bounded variation ($p = 1$) we only need to know its increments, while for an n-dimensional Brownian path, we need to define the second iterated integrals by specifying the rules on how to construct them. In general, we can think of a p-rough path as a path $X : [0,T] \to \mathbb{R}^n$ of finite p-variation, together with a set of rules on how to define the first $\lfloor p \rfloor$ iterated integrals. Once we know how to construct the first $\lfloor p \rfloor$, we know how to construct all of them.

Definition 2.3. *Let $X : [0,T] \to \mathbb{R}^n$ be a path. The set of all iterated integrals is called the* **signature of the path** *and is denoted by $S(X)$.*

Let's go back to equation (4). When X is a path of finite p-variation for $1 \leq p < 2$, then Y is a solution of (4) if

$$Y_t = y_0 + \int_0^t f(Y_s; \theta) \cdot dX_s, \ \forall t \in I. \tag{5}$$

However, when $p \geq 2$, it is not clear how to define the integral. We will see that in order to uniquely define the integral, we will need to know the first $\lfloor p \rfloor$ iterated integrals of X. That is, equation (4) can only make sense if we know the p-rough path $\mathbf{X}$ that corresponds to the path X with finite p-variation. We say that equation (4) is driven by the p-rough path $\mathbf{X}$.

Suppose that $\mathbf{Z}$ is a p-rough path in $\mathbb{R}^{\ell_1}$. We will first define $\int f(\mathbf{Z})d\mathbf{Z}$ where $f : \mathbb{R}^{\ell_1} \to \mathrm{L}(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2})$. We will assume that f is a polynomial of degree q – however, it is possible to define the integral for any $f \in \mathrm{Lip}(\gamma - 1)$ for $\gamma > p$ (see [3]). Since f is a polynomial, its Taylor expansion will be a finite sum:

$$f(z_2) = \sum_{k=0}^{q} f_k(z_1)\frac{(z_2 - z_1)^{\otimes k}}{k!}, \quad \forall z_1, z_2 \in \mathbb{R}^{\ell_1}$$

where $f_0 = f$ and $f_k : \mathbb{R}^{\ell_1} \to \mathrm{L}({\mathbb{R}^{\ell_1}}^{\otimes k}, \mathrm{L}(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2}))$ and for all $z \in \mathbb{R}^{\ell_1}$, $f_k(z)$ is a symmetric k-linear mapping from $\mathbb{R}^{\ell_1}$ to $\mathrm{L}(\mathbb{R}^{\ell_1}, \mathbb{R}^{\ell_2})$, for $k \geq 1$. If $\mathbf{Z}$ is a 1-rough path, then

$$f(Z_u) = \sum_{k=0}^{q} f_k(Z_s)\mathbf{Z}_{s,u}^k, \ \forall (s,u) \in \Delta_T$$

and

$$\tilde{\mathbf{Z}}_{s,t} := \int_s^t f(Z_u)dZ_u = \sum_{k=0}^{q} f_k(Z_s)\mathbf{Z}_{s,t}^{k+1} \ \forall (s,t) \in \Delta_T.$$

Note that $\tilde{\mathbf{Z}} := \int f(\mathbf{Z})d\mathbf{Z}$ defined above is also a 1-rough path in $\mathbb{R}^{\ell_2}$ and thus, its higher iterated integrals are uniquely defined. It is possible to show that the mapping $\int f : \Omega_1(\mathbb{R}^{\ell_1}) \to \Omega_1(\mathbb{R}^{\ell_2})$ sending $\mathbf{Z}$ to $\tilde{\mathbf{Z}}$ is continuous in the p-variation topology.

Remark 2.4. *We say that a sequence $\mathbf{Z}(\mathbf{r})$ of p-rough paths converges to a p-rough path $\mathbf{Z}$ in p-variation topology if there exists an $M \in \mathbb{R}$ and a sequence $a(r)$ converging to zero when $r \to \infty$, such that*

$$\begin{aligned} \|\mathbf{Z}(\mathbf{r})_{s,t}^i\|, \ \|\mathbf{Z}_{s,t}^i\| &\leq (M(t-s))^{\frac{i}{p}}, \text{ and} \\ \|\mathbf{Z}(\mathbf{r})_{s,t}^i - \mathbf{Z}_{s,t}^i\| &\leq a(r)(M(t-s))^{\frac{i}{p}} \end{aligned}$$

for $i = 1, \ldots, \lfloor p \rfloor$ and $(s,t) \in \Delta_T$. Note that this is not exactly equivalent to convergence in d_p: while convergence in d_p implies convergence in the p-variation topology, the opposite is not exactly true. Convergence in the p-variation topology implies that there is a subsequence that converges in d_p.

Now suppose that $\mathbf{Z}$ is a geometric p-rough path, i.e. $\mathbf{Z} \in G\Omega_p(\mathbb{R}^{\ell_1})$. By definition, there exists a sequence $\mathbf{Z}(\mathbf{r}) \in \Omega_1(\mathbb{R}^{\ell_1})$ such that $d_p(\mathbf{Z}(\mathbf{r}), \mathbf{Z}) \to 0$ as $r \to \infty$. We define $\tilde{\mathbf{Z}} := \int f(\mathbf{Z})d\mathbf{Z}$ as the limit of $\tilde{\mathbf{Z}}(\mathbf{r}) := \int f(\mathbf{Z}(\mathbf{r}))d\mathbf{Z}(\mathbf{r})$ with respect to d_p – this is will also be a geometric p-rough path. In other words, the continuous map $\int f$ can be extended to a continuous map from $G\Omega_p(\mathbb{R}^{\ell_1})$ to $G\Omega_p(\mathbb{R}^{\ell_2})$, which are the closures of $\Omega_1(\mathbb{R}^{\ell_1})$ and $\Omega_1(\mathbb{R}^{\ell_2})$ respectively (see Theorem 4.12, [3]).

We can now give the precise meaning of the solution of (4):

Definition 2.5. *Consider* $\mathbf{X} \in G\Omega_p(\mathbb{R}^n)$ *and* $y_0 \in \mathbb{R}^m$. *Set* $f_{y_0}(\cdot) := f(\cdot + y_0)$. *Define* $h : \mathbb{R}^n \oplus \mathbb{R}^m \to \mathrm{End}(\mathbb{R}^n \oplus \mathbb{R}^m)$ *by*

$$h(x,y) := \begin{pmatrix} I_{n\times n} & \mathbf{0}_{n\times m} \\ f_{y_0}(y) & \mathbf{0}_{m\times m} \end{pmatrix}. \tag{6}$$

We call $\mathbf{Z} \in G\Omega_p(\mathbb{R}^n \oplus \mathbb{R}^m)$ *a* **solution** *of (4) if the following two conditions hold:*

(i) $\mathbf{Z} = \int h(\mathbf{Z}) d\mathbf{Z}$.

(ii) $\pi_{\mathbb{R}^n}(\mathbf{Z}) = \mathbf{X}$, *where by* $\pi_{\mathbb{R}^n}$ *we denote the projection of* $\mathbf{Z}$ *to* $\mathbb{R}^n$.

As in the case of ordinary differential equations ($p = 1$), it is possible to construct the solution using Picard iterations: we define $\mathbf{Z}(\mathbf{0}) := (\mathbf{X}, \mathbf{e})$, where by $\mathbf{e}$ we denote the trivial rough path $\mathbf{e} = (1, \mathbf{0}_{\mathbb{R}^n}, \mathbf{0}_{\mathbb{R}^{n\otimes 2}}, \dots)$. Then, for every $r \geq 1$, we define $\mathbf{Z}(\mathbf{r}) = \int h(\mathbf{Z}(\mathbf{r-1})) d\mathbf{Z}(\mathbf{r-1})$. The following theorem, called the "Universal Limit Theorem", gives the conditions for the existence and uniqueness of the solution to (4). The theorem holds for any $f \in \mathrm{Lip}(\gamma)$ for $\gamma > p$ but we will assume that f is a polynomial. The proof is based on the convergence of the Picard iterations.

Theorem 2.6 (Theorem 5.3, [3])**.** *Let* $p \geq 1$. *For all* $\mathbf{X} \in G\Omega_p(\mathbb{R}^n)$ *and all* $y_0 \in \mathbb{R}^m$, *equation (4) admits a unique solution* $\mathbf{Z} = (\mathbf{X}, \mathbf{Y}) \in G\Omega_p(\mathbb{R}^n \oplus \mathbb{R}^m)$, *in the sense of definition 2.5. This solution depends continuously on* $\mathbf{X}$ *and* y_0 *and the mapping* $I_f : G\Omega_p(\mathbb{R}^n) \to G\Omega_p(\mathbb{R}^m)$ *which sends* $(\mathbf{X}, y_0)$ *to* $\mathbf{Y}$ *is continuous in the p-variation topology.*

The rough path $\mathbf{Y}$ *is the limit of the sequence* $\mathbf{Y}(r)$, *where* $\mathbf{Y}(r)$ *is the projection of the* r^{th} *Picard iteration* $\mathbf{Z}(r)$ *to* $\mathbb{R}^m$. *For all* $\rho > 1$, *there exists* $T_\rho \in (0, T]$ *such that*

$$\|\mathbf{Y}(r)^i_{s,t} - \mathbf{Y}(r+1)^i_{s,t}\| \leq 2^i \rho^{-r} \frac{(M(t-s))^{\frac{i}{p}}}{\beta\left(\frac{i}{p}\right)!}, \ \forall (s,t) \in \Delta_{T_\rho}, \ \forall i = 0, \dots, \lfloor p \rfloor.$$

The constant T_ρ *depends only on* f *and* p.

2.2 The problem

We now describe the problem that we are going to study in the rest of the paper. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbf{X} : \Omega \to G\Omega_p(\mathbb{R}^n)$ a random variable, taking values in the space of geometric p-rough paths endowed with the p-variation topology. For each $\omega \in \Omega$, the rough path $\mathbf{X}(\omega)$ drives the following differential equation

$$dY_t(\omega) = f(Y_t(\omega); \theta) \cdot dX_t(\omega), Y_0 = y_0 \tag{7}$$

where $\theta \in \Theta \subseteq \mathbb{R}^d$, Θ being the parameter space and for each $\theta \in \Theta$. As before, $f : \mathbb{R}^m \times \Theta \to \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $f_\theta(y) := f(y; \theta)$ is a polynomial on y for each $\theta \in \Theta$. According to theorem 2.6, we can think of equation (7) as a map

$$I_{f_\theta, y_0} : G\Omega_p(\mathbb{R}^n) \to G\Omega_p(\mathbb{R}^m), \tag{8}$$

sending a geometric p-rough path $\mathbf{X}$ to a geometric p-rough path $\mathbf{Y}$ and is continuous with respect to the p-variation topology. Consequently,

$$\mathbf{Y} := I_{f_\theta, y_0} \circ \mathbf{X} : \Omega \to G\Omega_p(\mathbb{R}^m)$$

is also a random variable, taking values in $G\Omega_p(\mathbb{R}^m)$ and if $\mathbb{P}^T$ is the distribution of $\mathbf{X}_{0,T}$, the distribution of $\mathbf{Y}_{0,T}$ will be $\mathbb{Q}_\theta^T = \mathbb{P}^T \circ I_{f_\theta, y_0}^{-1}$.

Suppose that the know the **expected signature** of $\mathbf{X}$ at $[0, T]$, i.e. we know

$$\mathbb{E}\left(\mathbf{X}_{0,T}^{(i_1,\ldots,i_k)}\right) := \mathbb{E}\left(\int \cdots \int_{0<u_1<\cdots<u_k<T} dX_{u_1}^{(i_1)} \cdots dX_{u_k}^{(i_k)}\right),$$

for all $i_j \in \{1, \ldots, n\}$ where $j = 1, \ldots, k$ and $k \geq 1$. Our goal will be to estimate θ, given several realizations of $\mathbf{Y}_{0,T}$, i.e. $\{\mathbf{Y}_{0,T}(\omega_i)\}_{i=1}^N$.

3 Method

In order to estimate θ, we are going to use a method that is similar to the "Method of Moments". The idea is simple: we will try to (partially) match the empirical signature of the observed p-rough path with the theoretical one, which is a function of the unknown parameters. Remember that the data we have available is several realizations of the p-rough path $\mathbf{Y}_{0,T}$ described in section 2.2. To make this more precise, let us introduce some notation: let

$$E^\tau(\theta) := \mathbb{E}_\theta(\mathbf{Y}_{0,T}^\tau)$$

be the *theoretical signature* corresponding to parameter value θ and word τ and

$$M_N^\tau := \frac{1}{N} \sum_{i=1}^{N} \mathbf{Y}_{0,T}^\tau(\omega_i) \tag{9}$$

be the *empirical signature*, which is a Monte Carlo approximation of the actual one. The word τ is constructed from the alphabet $\{1, \ldots, m\}$, i.e. $\tau \in W_m$ where $W_m := \bigcup_{k \geq 0} \{1, \ldots, m\}^k$. The idea is to find $\hat{\theta}$ such that

$$E^\tau(\hat{\theta}) = M_N^\tau, \ \forall \tau \in V \subset W_m$$

for some choice of a set of words V. Then, $\hat{\theta}$ will be our estimate.

Remark 3.1. *When $m = 1$, the expected signature of $\mathbf{Y}$ is equivalent to its moments, since*

$$\mathbf{Y}_{0,T}^{\overbrace{(1, \ldots, 1)}^{m}} = \frac{1}{m!}(Y_T - Y_0)^m.$$

Several questions arise:

(i) How can we get an analytic expression for $E^\tau(\theta)$ as a function of θ?

(ii) What is a good choice for V or, for $m = 1$, how do we choose which moments to match?

(iii) How good is $\hat{\theta}$ as an estimate?

We will try to answer these questions below.

3.1 Computing the Theoretical Signature

We need to get an analytic expression for the expected signature of the p-rough path $\mathbf{Y}$ at $(0, T)$, where $\mathbf{Y}$ is the solution of (7) in the sense described above. We are given the expected signature of the p-rough path $\mathbf{X}$ which is driving the equation, again at $(0, T)$. Unfortunately, we need to make one more approximation since the solution $\mathbf{Y}$ will not usually be available: we will approximate the solution by the r^{th} Picard iteration $\mathbf{Y}(\mathbf{r})$, described in the Universal Limit Theorem (Theorem 2.6). Finally, we will approximate the expected signature of the solution corresponding to a word τ, $E^{\tau}(\theta)$, by the expected signature of the r^{th} Picard iteration at τ, which we will denote by $E_r^{\tau}(\theta)$:

$$E_r^{\tau}(\theta) := \mathbb{E}_{\theta}(\mathbf{Y}(\mathbf{r})_{0,T}^{\tau}). \tag{10}$$

The good news is that when f_{θ} is a polynomial of degree q on y, for any $q \in \mathbb{N}$, the r^{th} Picard iteration of the solution is a linear combination of iterated integrals of the driving force $\mathbf{X}$. More specifically, for any realization ω and any time interval $(s, t) \in \Delta_T$, we can write:

$$\mathbf{Y}(\mathbf{r})_{s,t}^{\tau} = \sum_{|\sigma| \leq |\tau| \frac{q^r - 1}{q-1}} \alpha_{r,\sigma}^{\tau}(y_0, s; \theta) \mathbf{X}_{s,t}^{\sigma}, \tag{11}$$

where $\alpha_{r,\sigma}^{\tau}(y; \theta)$ is a polynomial in y of degree q^r and $| \cdot |$ gives the length of a word.

We will prove this claim, first for $p = 1$ and then for any $p \geq 1$ by taking limits with respect to d_p. We will need the following lemma.

Lemma 3.2. *Suppose that $\mathbf{X} \in G\Omega_1(\mathbb{R}^n)$, $\mathbf{Y} \in G\Omega_1(\mathbb{R}^m)$ and it is possible to write*

$$\mathbf{Y}_{s,t}^{(j)} = \sum_{\sigma \in W_n,\ q_1 \leq |\sigma| \leq q_2} \alpha_{\sigma}^{(j)}(y_s) \mathbf{X}_{s,t}^{\sigma}, \ \ \forall (s,t) \in \Delta_T \text{ and } \forall j = 1, \ldots, m \tag{12}$$

where $\alpha_{\sigma}^{(j)} : \mathbb{R}^m \to \mathrm{L}(\mathbb{R}, \mathbb{R})$ is a polynomial of degree q and $q, q_1, q_2 \in \mathbb{N}$ and $q_1 \geq 1$. Then,

$$\mathbf{Y}_{s,t}^{\tau} = \sum_{\sigma \in W_n,\ |\tau| q_1 \leq |\sigma| \leq |\tau| q_2} \alpha_{\sigma}^{\tau}(y_s) \mathbf{X}_{s,t}^{\sigma}, \tag{13}$$

for all $(s,t) \in \Delta_T$ and $\tau \in W_m$. $\alpha_{\sigma}^{\tau} : \mathbb{R}^m \to \mathrm{L}(\mathbb{R}, \mathbb{R})$ are polynomials of degree $\leq q|\tau|$.

Proof. We will prove (13) by induction on $|\tau|$, i.e. the length of the word. By hypothesis, it is true when $|\tau| = 1$. Suppose that it is true for any $\tau \in W_m$ such that $|\tau| = k \geq 1$. First, note that from (12), we get that

$$dY_u^{(j)} = \sum_{\sigma \in W_n,\ q_1 \leq |\sigma| \leq q_2} \alpha_{\sigma}^{(j)}(y_s) \mathbf{X}_{s,u}^{\sigma-} dX_u^{\sigma_\ell}, \ \ \forall u \in [s,t]$$

where $\sigma-$ is the word σ without the last letter and σ_ℓ is the last letter. For example, if $\sigma = (i_1, \dots, i_{b-1}, i_b)$, then $\sigma- = (i_1, \dots, i_{b-1})$ and $\sigma_\ell = i_b$. Note that this cannot be defined when σ is the empty word $\emptyset$ ($b = 0$). Now suppose that $|\tau| = k+1$, so $\tau = (j_1, \dots, j_k, j_{k+1})$ for some $j_1, \dots, j_{k+1} \in \{1, \dots, m\}$. Then

$$\begin{aligned}
\mathbf{Y}^\tau_{s,t} &= \int_s^t \mathbf{Y}^{\tau-}_{s,u} dY^{(j_{k+1})}_u = \\
&= \int_s^t \left(\sum_{kq_1 \le |\sigma_1| \le kq_2} \alpha^{\tau-}_{\sigma_1}(y_s) \mathbf{X}^{\sigma_1}_{s,u} \right) \sum_{q_1 \le |\sigma_2| \le q_2} \alpha^{(j_{k+1})}_{\sigma_2}(y_s) \mathbf{X}^{\sigma_2 -}_{s,u} dX^{\sigma_{2\ell}}_u = \\
&= \sum_{kq_1 \le |\sigma_1| \le kq_2,\ q_1 \le |\sigma_2| \le q_2} \left(\alpha^{\tau-}_{\sigma_1}(y_s) \alpha^{(j_{k+1})}_{\sigma_2}(y_s) \right) \int_s^t \mathbf{X}^{\sigma_1}_{s,u} \mathbf{X}^{\sigma_2 -}_{s,u} dX^{\sigma_{2\ell}}_u .
\end{aligned}$$

Now we use the fact that for any geometric rough path $\mathbf{X}$ and any $(s,u) \in \Delta_T$, we can write

$$\mathbf{X}^{\sigma_1}_{s,u} \mathbf{X}^{\sigma_2 -}_{s,u} = \sum_{\sigma \in \sigma_1 \sqcup (\sigma_2 -)} \mathbf{X}^\sigma_{s,u}, \tag{14}$$

where $\sigma_1 \sqcup (\sigma_2-)$ is the **shuffle product** between the words σ_1 and σ_2-, i.e. it is the set of all words that we can create by mixing up the letters of σ_1 and σ_2- without changing the order of letters within each word. For example, $(1,2) \sqcup (1) = \{(1,2,2), (1,2,2), (2,1,2)\}$ (see [3]). Applying (14) above, we get

$$\mathbf{Y}^\tau_{s,t} = \sum_{\sigma \in W_n,\ (k+1)q_1 \le |\sigma| \le (k+1)q_2} \alpha^\tau_\sigma(y_s) \mathbf{X}^\sigma_{s,t},$$

where

$$\alpha^\tau_\sigma(y_s) = \sum_{(\sigma_1 \sqcup \sigma_2 -) \ni \sigma-,\ \sigma_\ell = \sigma_{2_\ell}} \alpha^{\tau-}_{\sigma_1}(y_s) \alpha^{\tau_\ell}_{\sigma_2}(y_s)$$

is a polynomial of degree $\le kq + q = (k+1)q$. Note that the above sum is over all $\sigma_1, \sigma_2 \in W_n$ such that $kq_1 \le |\sigma_1| \le kq_2$ and $q_1 \le |\sigma_1| \le q_2$ □

We now prove (11) for $p = 1$.

Lemma 3.3. *Suppose that $\mathbf{X} \in G\Omega_1(\mathbb{R}^n)$ is driving system (4), where $f : \mathbb{R}^m \to \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ is a polynomial of degree q. Let $\mathbf{Y}(\mathbf{r})$ be the projection of the r^{th} Picard iteration $\mathbf{Z}(\mathbf{r})$ to $\mathbb{R}^m$, as described above. Then, $\mathbf{Y}(\mathbf{r}) \in G\Omega_1(\mathbb{R}^m)$ and it satisfies*

$$\mathbf{Y}(\mathbf{r})^\tau_{s,t} = \sum_{|\sigma| \le |\tau| \frac{q^r - 1}{q-1}} \alpha^\tau_{r,\sigma}(y_0, s) \mathbf{X}^\sigma_{s,t}, \tag{15}$$

for all $(s,t) \in \Delta_T$ and $\tau \in W_m$. $\alpha^\tau_{r,\sigma}(y,s)$ is a polynomial of degree $\le |\tau| q^r$ in y.

Proof. For every $r \ge 0$, $\mathbf{Z}(\mathbf{r}) \in G\Omega_1(\mathbb{R}^{n+m})$ since $\mathbf{Z}(\mathbf{0}) := (\mathbf{X}, \mathbf{e})$, $\mathbf{X} \in G\Omega_1(\mathbb{R}^n)$ and integrals preserve the roughness of the integrator. So, $\mathbf{Y}(\mathbf{r}) \in G\Omega_1(\mathbb{R}^m)$. We will prove the claim by induction on r.

For $r = 0$, $\mathbf{Y}(\mathbf{0}) = \mathbf{e}$ and thus (15) becomes

$$\mathbf{Y}(\mathbf{0})^{\tau}_{s,t} = \alpha^{\tau}_{0,\emptyset}(y_0, s)$$

and it is true for $\alpha^{\emptyset}_{0,\emptyset} \equiv 1$ and $\alpha^{\tau}_{0,\emptyset} \equiv 0$ for every $\tau \in W_m$ such that $|\tau| > 0$.

Now suppose it is true for some $r \geq 0$. Remember that $\mathbf{Z}(\mathbf{r}) = (\mathbf{X}, \mathbf{Y}(\mathbf{r}))$ and that $\mathbf{Z}(\mathbf{r+1})$ is defined by

$$\mathbf{Z}(\mathbf{r+1}) = \int h(\mathbf{Z}(\mathbf{r}))d\mathbf{Z}(\mathbf{r})$$

where h is defined in (6) and $f_{y_0}(y) = f(y_0 + y)$. Since f is a polynomial of degree q, h is also a polynomial of degree q and, thus, it is possible to write

$$h(z_2) = \sum_{k=0}^{q} h_k(z_1)\frac{(z_2 - z_1)^{\otimes k}}{k!}, \quad \forall z_1, z_2 \in \mathbb{R}^{\ell}, \tag{16}$$

where $\ell = n + m$. Then, the integral is defined to be

$$\mathbf{Z}(\mathbf{r+1})_{s,t} := \int_s^t h(\mathbf{Z}(\mathbf{r}))d\mathbf{Z}(\mathbf{r}) = \sum_{k=0}^{q} h_k\left(Z(r)_s\right)\mathbf{Z}(\mathbf{r})^{k+1}_{s,t} \; \forall (s,t) \in \Delta_T.$$

Let's take a closer look at functions $h_k : \mathbb{R}^{\ell} \to \mathrm{L}\left(\mathbb{R}^{\ell \otimes k}, \mathrm{L}(\mathbb{R}^{\ell}, \mathbb{R}^{\ell})\right)$. Since (16) is the Taylor expansion for polynomial h, h_k is the k^{th} derivative of h. So, for every work $\beta \in W_{\ell}$ such that $|\beta| = k$ and every $z = (x, y) \in \mathbb{R}^{\ell}$, $(h_k(z))^{\beta} = \partial_{\beta} h(z) \in \mathrm{L}(\mathbb{R}^{\ell}, \mathbb{R}^{\ell})$. By definition, h is independent of x and thus the derivative will always be zero if β contains any letters in $\{1, \ldots, n\}$.

Remember that $\mathbf{Y}(\mathbf{r+1})$ is the projection of $\mathbf{Z}(\mathbf{r+1})$ onto $\mathbb{R}^m$. So, for each $j \in \{1, \ldots, m\}$,

$$\begin{aligned}
\mathbf{Y}(\mathbf{r+1})^{(j)}_{s,t} &= \mathbf{Z}(\mathbf{r+1})^{(n+j)}_{s,t} = \sum_{k=0}^{q}\left(h_k\left(Z(r)_s\right)\mathbf{Z}(\mathbf{r})^{k+1}_{s,t}\right)^{(n+j)} = \\
&= \sum_{k=0}^{q}\sum_{i=1}^{\ell}\sum_{\tau \in W_m, |\tau| = k} \partial_{\tau+n} h_{n+j,i}\left(Z(r)_s\right)\mathbf{Z}(\mathbf{r})^{(\tau+n,i)}_{s,t} = \\
&= \sum_{k=0}^{q}\sum_{i=1}^{n}\sum_{\tau \in W_m, |\tau| = k} \partial_{\tau} f_{j,i}\left(y_0 + Y(r)_s\right)\mathbf{Z}(\mathbf{r})^{(\tau+n,i)}_{s,t}.
\end{aligned}$$

By the induction hypothesis, we know that for every $\tau \in W_m$,

$$\mathbf{Z}(\mathbf{r})^{\tau+n}_{s,t} = \mathbf{Y}(\mathbf{r})^{\tau}_{s,t} = \sum_{|\sigma| \leq |\tau|\frac{q^r-1}{q-1}} \alpha^{\tau}_{r,\sigma}(y_0, s)\mathbf{X}^{\sigma}_{s,t}$$

and thus, for every $i = 1, \ldots, n$,

$$\mathbf{Z}(\mathbf{r})^{(\tau+n,i)}_{s,t} = \sum_{|\sigma| \leq |\tau|\frac{q^r-1}{q-1}} \alpha^{\tau}_{r,\sigma}(y_0, s)\mathbf{X}^{(\sigma,i)}_{s,t}.$$

Putting this back to the equation above, we get

$$\mathbf{Y}(\mathbf{r}+\mathbf{1})^{(j)}_{s,t} = \sum_{i=1}^{n} \sum_{|\tau|\leq q} \partial_\tau f_{j,i}\left(y_0 + Y(r)_s\right) \sum_{|\sigma|\leq|\tau|\frac{q^r-1}{q-1}} \alpha^\tau_{r,\sigma}(y_0,s)\mathbf{X}^{(\sigma,i)}_{s,t}$$

and by re-organizing the sums, we get

$$\mathbf{Y}(\mathbf{r}+\mathbf{1})^{(j)}_{s,t} = \sum_{|\sigma|\leq q\frac{q^r-1}{q-1}+1=\frac{q^{r+1}-1}{q-1}} \alpha^{(j)}_{r+1,\sigma}(y_0,s)\mathbf{X}^{\sigma}_{s,t},$$

where $\alpha^{(j)}_{r+1,\emptyset} \equiv 0$ and for every $\sigma \in W_n - \emptyset$,

$$\alpha^{(j)}_{r+1,\sigma}(y_0,s) = \sum_{\frac{|\sigma-|(q-1)}{q^r-1}\leq|\tau|\leq q} \partial_\tau f_{j,\sigma_\ell}(y_0 + Y(r)_s)\alpha^\tau_{r,\sigma-}(y_0,s).$$

If $\alpha^\tau_{r,\sigma}$ are polynomials of degree $\leq |\tau|q^r$, then $\alpha^{(j)}_{r,\sigma}$ are polynomials of degree $\leq q^r$. The result follow by applying lemma 3.2. Notice that (in the notation of lemma 3.2) $q_1 \geq 1$ since $\alpha^{(j)}_{r+1,\emptyset} \equiv 0$. $\square$

We will now prove (11) for any $p \geq 1$.

Theorem 3.4. *The result of lemma 3.3 still holds when $\mathbf{X} \in G\Omega_p(\mathbb{R}^n)$, for any $p \geq 1$.*

Proof. Since $\mathbf{X} \in G\Omega_p(\mathbb{R}^n)$, there exists a sequence $\{\mathbf{X}(\mathbf{k})\}_{k\geq 0}$ in $G\Omega_1(\mathbb{R}^n)$, such that $\mathbf{X}(\mathbf{k}) \overset{k\to\infty}{\to} \mathbf{X}$ in the p-variation topology. We denote by $\mathbf{Z}(\mathbf{k},\mathbf{r})$ and $\mathbf{Z}(\mathbf{r})$ the r^{th} Picard iteration corresponding to equation (4) driven by $\mathbf{X}(\mathbf{k})$ and $\mathbf{X}$ respectively.

First, we show that $\mathbf{Z}(\mathbf{k},\mathbf{r}) \overset{k\to\infty}{\to} \mathbf{Z}(\mathbf{r})$ and consequently $\mathbf{Y}(\mathbf{k},\mathbf{r}) \overset{k\to\infty}{\to} \mathbf{Y}(\mathbf{r})$ in the p-variation topology, for every $r \geq 0$. It is clearly true for $r = 0$. Now suppose that it is true for some $r \geq 0$. By definition, $\mathbf{Z}(\mathbf{r}+\mathbf{1}) = \int h(\mathbf{Z}(\mathbf{r}))d\mathbf{Z}(\mathbf{r})$. Remember that the integral is defined as the limit in the p-variation topology of the integrals corresponding to a sequence of 1-rough paths that converge to $\mathbf{Z}(\mathbf{r})$ in the p-variation topology. By the induction hypothesis, this sequence can be $\mathbf{Z}(\mathbf{k},\mathbf{r})$. It follows that $\mathbf{Z}(\mathbf{k},\mathbf{r}+\mathbf{1}) = \int h(\mathbf{Z}(\mathbf{k},\mathbf{r}))d\mathbf{Z}(\mathbf{k},\mathbf{r})$ converges to $\mathbf{Z}(\mathbf{r}+\mathbf{1})$, which proves the claim. Convergence of the rough paths in p-variation topology implies convergence of each of the iterated integrals, i.e.

$$\mathbf{Y}(\mathbf{k},\mathbf{r})^\tau_{s,t} \overset{k\to\infty}{\to} \mathbf{Y}(\mathbf{r})^\tau_{s,t}$$

for all $r \geq 0$, $(s,t) \in \Delta_T$ and $\tau \in W_m$.

By lemma 3.3, since $\mathbf{X}(\mathbf{k}) \in G\Omega_1(\mathbb{R}^n)$ for every $k \geq 1$, we can write

$$\mathbf{Y}(\mathbf{k},\mathbf{r})^\tau_{s,t} = \sum_{|\sigma|\leq|\tau|\frac{q^r-1}{q-1}} \alpha^\tau_{r,\sigma}(y_0,s)\mathbf{X}(\mathbf{k})^\sigma_{s,t},$$

for every $\tau \in W_m$, $(s,t) \in \Delta_T$ and $k \geq 1$. Since $\mathbf{X}(\mathbf{k}) \overset{k\to\infty}{\to} \mathbf{X}$ in the p-variation topology and the sum is finite, it follows that

$$\mathbf{Y}(\mathbf{k},\mathbf{r})^{\tau}_{s,t} \overset{k\to\infty}{\to} \sum_{|\sigma|\leq|\tau|\frac{q^r-1}{q-1}} \alpha^{\tau}_{r,\sigma}(y_0,s)\mathbf{X}^{\sigma}_{s,t}.$$

The statement of the theorem follows. □

3.2 The Signature Matching Estimator

We can now give a precise definition of the estimator, which we will formally call the **Signature Matching Estimator** (SME): suppose that we are in the setting of the problem described in section 2.2 and M^{τ}_N and $E^{\tau}_r(\theta)$ are defined as in (9) and (10) respectively, for every $\tau \in W_m$. Let $V \subset W_m$ be a set of d words constructed from the alphabet $\{1,\ldots,m\}$. For each such V, we define the SME $\hat{\theta}^V_{r,N}$ as the solution to

$$E^{\tau}_r(\theta) = M^{\tau}_N,\ \forall\tau \in V. \tag{17}$$

This definition requires that (17) has a *unique* solution. This will not be true in general. Let $\mathcal{V}_r$ be the set of all V such that $E^{\tau}_r(\theta) = M,\ \forall\tau \in V$ has a unique solution for all $M \in S_\tau \subseteq \mathbb{R}$ where S_τ is the set of all possible values of M^{τ}_N, for any $N \geq 1$. We will assume the following:

Assumption 1 (Observability)**.** *The set $\mathcal{V}_r$ is non-empty and known (at least up to a non-empty subset).*

Then, $\hat{\theta}^V_{r,N}$ can be defined for every $V \in \mathcal{V}_r$.

Remark 3.5. *In order to achieve uniqueness of the estimator, we might need some extra information that we could get by looking at time correlations. We can fit this into our framework by considering scaled versions of (7) together with the original one: for example consider the equation*

$$\begin{aligned} dY_t(\omega) &= f(Y_t(\omega);\theta)\cdot dX_t(\omega), Y_0 = y_0 \\ dY(c)_t(\omega) &= f(Y(c)_t(\omega);\theta)\cdot dX_{ct}(\omega), Y(c)_0 = y_0 \end{aligned}$$

for some appropriate constant c. Then, $Y(c)_t = Y_{ct}$ and the expected signature at $[0,T]$ will also contain information about $\mathbb{E}\left(Y^{(j_1)}_T Y^{(j_2)}_{cT}\right)$ for any $j_1, j_2 = 1,\ldots,m$.

It is very difficult to say anything about the solutions of system (17), as it is very general. However, if we assume that f is also a polynomial in θ, then (17) becomes a system of polynomial equations. In the appendix, we study this system in more detail for the general case where θ are the coefficients of the Taylor expansion of $f(y;\theta)$ around y_0, which corresponds to the case were we know nothing about the polynomials f except their degree.

Note that one can also create a Generalized Signature Matching Estimator as the solution of

$$P_\alpha\left(E^{\tau}_r(\theta)\right) = P_\alpha\left(M^{\tau}_N\right),\ \text{for } \alpha \in A$$

where P_α are polynomials of (empirical or theoretical) expected values of iterated integrals corresponding to words τ and A an appropriate index set.

Remark 3.6. *In the case where y_t is a Markov process, the Generalized Moment Matching Estimator can be seen as a special case of the Generalized Signature Matching Estimator. In that case, the question of identifiability has been studied in detail (see [5]), but without considering the extra approximation of the theoretical moments by Picard iteration.*

3.3 Properties of the SME

It is possible to show that the SME defined as the solution of (17) will converge to the true value of the parameter and will have the asymptotic normality property. More precisely, the following holds:

Theorem 3.7. *Let $\hat{\theta}^V_{r,N}$ be the Signature Matching Estimator for the system described in section 2.2 and $V \in \mathcal{V}_r$. Assume that $f(y;\theta)$ is a polynomial of degree q with respect to y and twice differentiable with respect to θ and θ_0 is the 'true' parameter value, meaning that the distribution of the observed signature $\mathbf{Y}_{0,T}$ is $\mathbb{Q}^T_{\theta_0}$. Set*

$$D^V_r(\theta)_{i,\tau} = \frac{\partial}{\partial \theta_i} E^\tau_r(\theta) \quad \textit{and} \quad \Sigma_V(\theta_0)_{\tau,\tau'} = \mathrm{cov}\left(\mathbf{Y}^\tau_{0,T}, \mathbf{Y}^{\tau'}_{0,T}\right)$$

and assume that $\inf_{r>0,\theta\in\Theta} \|D^V_r(\theta)\| > 0$, i.e. $D^V_r(\theta)$ is uniformly non-degenerate with respect to r and θ. Then, for $r \propto \log N$ and T are sufficiently small,

$$\hat{\theta}^V_{r,N} \to \theta_0, \quad \textit{with probability } 1, \tag{18}$$

and

$$\sqrt{N}\Phi_V(\theta_0)^{-1}\left(\hat{\theta}^V_{r,N} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I) \tag{19}$$

as $N \to \infty$, where

$$\Phi_V(\theta_0) = D^V_r(\theta_0)^{-1}\Sigma_V(\theta_0).$$

Proof. By theorem 3.4 and the definition of $E^\tau_r(\theta)$,

$$E^\tau_r(\theta) = \sum_{|\sigma|\le|\tau|\frac{q^r-1}{q-1}} \alpha^\tau_{r,\sigma}(y_0;\theta)\mathbb{E}\left(\mathbf{X}^\sigma_{0,T}\right)$$

where functions $\alpha^\tau_{r,\sigma}(y_0;\theta)$ are constructed recursively, as in lemmas 3.2 and 3.3. Since f is twice differentiable with respect to θ, functions α and consequently E^τ_r will also be twice differentiable with respect to θ. Thus, we can write

$$E^\tau_r(\theta) - E^\tau_r(\theta_0) = D^V_r(\tilde{\theta})_{\cdot,\tau}(\theta - \theta_0),\ \forall \theta \in \Theta \subseteq \mathbb{R}^d$$

for some $\tilde{\theta}$ within a ball of center θ_0 and radius $\|\theta - \theta_0\|$ and the function $D^V_r(\theta)$ is continuous. By inverting D^V_r and for $\theta = \hat{\theta}^V_{r,N}$, we get

$$\left(\hat{\theta}^V_{r,N} - \theta_0\right) = D^V_r(\tilde{\theta}^V_{r,N})^{-1}\left(E^V_r(\hat{\theta}^V_{r,N}) - E^V_r(\theta_0)\right) \tag{20}$$

where $E_r^V(\theta) = \{E_r^\tau(\theta)\}_{\tau \in V}$. By definition

$$E_r^V(\hat{\theta}_{r,N}^V) = \{M_N^\tau\}_{\tau \in V} = \{\frac{1}{N}\sum_{i=1}^N \mathbf{Y}_{0,T}^\tau(\omega_i)\}_{\tau \in V} \tag{21}$$

where $\mathbf{Y}_{0,T}(\omega_i)$ are independent realizations of the random variable $\mathbf{Y}_{0,T}$. Suppose that T is small enough, so that the above Monte-Carlo approximation satisfies both the Law of Large Numbers and the Central Limit Theorem, i.e. the covariance matrix satisfies $0 < \|\Sigma_V(\theta_0)\| < \infty$. Then, for $N \to \infty$

$$E_r^\tau(\hat{\theta}_{r,N}^V) \to E^\tau(\theta_0) = \mathbb{E}\left(\mathbf{Y}_{0,T}^\tau\right), \quad \forall \tau \in V$$

with probability 1. Note that the convergence does not depend on r. Also, for $r \to \infty$

$$E_r^\tau(\theta_0) \to E^\tau(\theta_0)$$

as a result of theorem 2.6. Thus, for $r \propto \log N$

$$\|E_r^\tau(\hat{\theta}_{r,N}^V) - E_r^\tau(\theta)\| \to 0, \quad \text{with probability 1.}$$

Combining this with (20) and the uniform non-degeneracy of D_r^V, we get (18). From (18) and the continuity and uniform non-degeneracy of D_r^V, we conclude that

$$D_r^V(\theta_0) D_r^V(\tilde{\theta}_{r,N}^V)^{-1} \to I, \quad \text{with probability 1}$$

provided that T is small enough, so that $E^V(\theta_0) < \infty$. Now, since

$$\Phi_V(\theta_0)^{-1}\left(\hat{\theta}_{r,N}^V - \theta_0\right) = \Sigma_V(\theta_0)^{-1}\left(D_r^V(\theta_0) D_r^V(\tilde{\theta}_{r,N}^V)^{-1}\right)\left(E_r^V(\hat{\theta}_{r,N}^V) - E_r^V(\theta_0)\right)$$

to prove (19) it is sufficient to prove that

$$\sqrt{N}\Sigma_V(\theta_0)^{-1}\left(E_r^V(\hat{\theta}_{r,N}^V) - E_r^V(\theta_0)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I)$$

It follows directly from (21) that

$$\sqrt{N}\Sigma_V(\theta_0)^{-1}\left(E_r^V(\hat{\theta}_{r,N}^V) - E^V(\theta_0)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I).$$

It remains to show that

$$\sqrt{N}\Sigma_V(\theta_0)^{-1}\left(E_r^V(\theta_0) - E^V(\theta_0)\right) \to 0.$$

It follows from theorem 2.6 that

$$\|E_r^V(\theta_0) - E^V(\theta_0)\| \le C\rho^{-r}$$

for any $\rho > 1$ and sufficiently small T. The constant C depends on V, p and T. Suppose that $r = a\log N$ for some $a > 0$ and choose $\rho > \exp\left(\frac{1}{2c}\right)$. Then

$$\sqrt{N}\|\left(E_r^V(\theta_0) - E^V(\theta_0)\right)\| \le CN^{\left(\frac{1}{2} - c\log\rho\right)}$$

which proves the claim. $\square$

Remark 3.8. *We have now completed the discussion of the questions set in remark 3.1: we provided a way for getting an analytic expression for an approximation of* $E^\tau(\theta)$*. Also, the asymptotic variance of the estimator can be used to compare different choices of* V *and to assess the quality of the estimator.*

In the case of diffusions and the GMM estimator, a discussion on how to optimally choose which moments to work on can be found in [6].

References

[1] J. Li, P.G. Kevrekidis, C.W. Gear and I.G. Kevrekidis. Deciding the nature of the coarse equation through microscopic simulations: the baby-bathwater scheme. *SIAM Rev.* 49(3):469–487, 2007.

[2] A. Papavasiliou, G.A. Pavliotis and A.M. Stuart. Maximum Likelihood Drift Estimation for Multiscale Diffusions. *Stochastic Processes Applications* (to appear).

[3] T.J. Lyons, M.J. Caruna and T. Lévy. *Differential Equations Driven by Rough Paths*. In *Ecole d'été de Probabilités de Saint-Flour XXXIV* (Editor: J. Picard), Springer, Berlin 2007.

[4] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Application of Mathematics 23. Springer-Verlag, Berlin 1999.

[5] L. P. Hansen and J. A. Scheinkman. Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica* 63(4):767-804, 1995.

[6] Y. Aït-Sahalia and P. A. Mykland. An Analysis of Hansen-Scheinkman Moment Estimators for Discretely and Randomly Sampled Diffusions. *Journal of Econometrics* 144:1–26, 2008.