Efficient drift parameter estimation for ergodic solutions of backward SDEs [1]

Teppei Ogihara

The University of Tokyo

September 26, 2024
Fall School Time Series, Random Fields and beyond

Joint work with Mitja Stadje (Ulm University)

[1]Ogihara, T. and Stadje, M.: Efficient drift parameter estimation for ergodic solutions of backward SDEs, Scandinavian Journal of Statistics, 51(3), 1181–1205. (2024)

Overview

- Consider a parameter estimation problem for an m-dimensional stochastic differential equation (SDE) model:

$$Y_t = Y_0 + \int_0^t \psi(X_s, Y_s, V_s V_s^\top, \theta_0) ds + \int_0^t V_s dW_s, \qquad (1)$$

W_t: r-dimensional Brownian motion (with independent increments $W_t - W_s \sim N(0, (t-s)I_r))$

Overview

- Consider a parameter estimation problem for an m-dimensional stochastic differential equation (SDE) model:

$$Y_t = Y_0 + \int_0^t \psi(X_s, Y_s, V_s V_s^\top, \theta_0) ds + \int_0^t V_s dW_s, \qquad (1)$$

  W_t: r-dimensional Brownian motion (with independent increments $W_t - W_s \sim N(0, (t-s)I_r)$)

- We observe $\{(X_{kh_n}, Y_{kh_n})\}_{k=0}^n$ with $h_n \to 0$, $nh_n \to \infty$, and $nh_n^2 \to 0$.

Overview

- Consider a parameter estimation problem for an m-dimensional stochastic differential equation (SDE) model:

$$Y_t = Y_0 + \int_0^t \psi(X_s, Y_s, V_s V_s^\top, \theta_0) ds + \int_0^t V_s dW_s, \qquad (1)$$

  W_t: r-dimensional Brownian motion (with independent increments $W_t - W_s \sim N(0, (t-s)I_r)$)
- We observe $\{(X_{kh_n}, Y_{kh_n})\}_{k=0}^n$ with $h_n \to 0$, $nh_n \to \infty$, and $nh_n^2 \to 0$.
- Unlike previous studies, we estimate the drift parameter θ_0 when the diffusion coefficient V_t is unknown. This model is a type of backward SDE, and in many situations, V_t is unknown and unobserved.

When V_t is parametric

We first consider cases where V_t can be written as

$$V_t = b(Y_t, \sigma_0), \quad \psi(x, y, z, \theta) = a(y, \theta)$$

using $\mathbb{R}^m \otimes \mathbb{R}^r$-valued function $b(y, \sigma)$ and $\mathbb{R}^m$-valued function $a(y, \theta)$, and we observe $\{Y_{kh_n}\}_{k=0}^n$.

When V_t is parametric

We first consider cases where V_t can be written as

$$V_t = b(Y_t, \sigma_0), \quad \psi(x, y, z, \theta) = a(y, \theta)$$

using $\mathbb{R}^m \otimes \mathbb{R}^r$-valued function $b(y, \sigma)$ and $\mathbb{R}^m$-valued function $a(y, \theta)$, and we observe $\{Y_{kh_n}\}_{k=0}^{n}$.

Let $t_k = kh_n$, then

$$\begin{aligned} \Delta_k Y &:= Y_{t_k} - Y_{t_{k-1}} \\ &\approx a_k(\theta_0)h_n + b_k(\sigma_0)(W_{t_k} - W_{t_{k-1}}) \\ &\sim N(a_k(\theta_0)h_n, b_k b_k^\top(\sigma_0)h_n), \quad \text{(conditional on } Y_{t_{k-1}}) \end{aligned}$$

Here $\top$ denotes matrix transpose, $a_k(\theta) = a(Y_{t_{k-1}}, \theta)$, $b_k(\sigma) = b(Y_{t_{k-1}}, \sigma)$.

Maximum Likelihood-type Estimation

Therefore, we can set up a quasi-log-likelihood function as follows:

$$H_n^0(\sigma,\theta) = -\frac{1}{2}\sum_{k=1}^{n}\left\{\bar{Y}_k(\theta)^\top (b_k b_k^\top(\sigma) h_n)^{-1}\bar{Y}_k(\theta) + \log\det(b_k b_k(\sigma)^\top)\right\}.$$

where $\bar{Y}_k(\theta) = \Delta Y_k - a_k(\theta)h_n$. (Local Gaussian approximation)

Maximum Likelihood-type Estimation

Therefore, we can set up a quasi-log-likelihood function as follows:

$$H_n^0(\sigma,\theta) = -\frac{1}{2}\sum_{k=1}^{n}\left\{\bar{Y}_k(\theta)^\top (b_k b_k^\top(\sigma)h_n)^{-1}\bar{Y}_k(\theta) + \log\det(b_k b_k(\sigma)^\top)\right\}.$$

where $\bar{Y}_k(\theta) = \Delta Y_k - a_k(\theta)h_n$. (Local Gaussian approximation)

Define the maximum likelihood-type estimator as:

$$(\hat{\sigma}_n^0, \hat{\theta}_n^0) \in \mathrm{argmax}_{\sigma,\theta} H_n^0(\sigma,\theta).$$

Maximum Likelihood-type Estimation

Therefore, we can set up a quasi-log-likelihood function as follows:

$$H_n^0(\sigma,\theta) = -\frac{1}{2}\sum_{k=1}^{n}\left\{\bar{Y}_k(\theta)^\top (b_k b_k^\top(\sigma)h_n)^{-1}\bar{Y}_k(\theta) + \log\det(b_k b_k(\sigma)^\top)\right\}.$$

where $\bar{Y}_k(\theta) = \Delta Y_k - a_k(\theta)h_n$. (Local Gaussian approximation)

Define the maximum likelihood-type estimator as:

$$(\hat{\sigma}_n^0, \hat{\theta}_n^0) \in \mathrm{argmax}_{\sigma,\theta} H_n^0(\sigma,\theta).$$

Theorem 1 (Kessler (1997), Yoshida (2011))

Under appropriate conditions on the diffusion coefficients a and b (smoothness, non-degeneracy, etc.), there exists a p.d. matrix Γ_0 such that as $n \to \infty$,

$$(\sqrt{n}(\hat{\sigma}_n^0 - \sigma_0), \sqrt{nh_n}(\hat{\theta}_n^0 - \theta_0)) \xrightarrow{d} N(0, \Gamma_0^{-1}).$$

When V_t is nonparametric

Consider the case where we don't assume a parametric model for V_t, and ψ includes $V_t V_t^\top$:

$$Y_t = Y_0 + \int_0^t \psi(X_s, Y_s, V_s V_s^\top, \theta_0) ds + \int_0^t V_s dW_s.$$

Consider discrete observations $\{(X_{kh_n}, Y_{kh_n})\}$, $k = 0, ..., n$. Since V_t is unknown and H_n^0 cannot be calculated, we approximate V_t using observational data. This model is a type of backward SDE.

When V_t is nonparametric

Consider the case where we don't assume a parametric model for V_t, and ψ includes $V_t V_t^\top$:

$$Y_t = Y_0 + \int_0^t \psi(X_s, Y_s, V_s V_s^\top, \theta_0) ds + \int_0^t V_s dW_s.$$

Consider discrete observations $\{(X_{kh_n}, Y_{kh_n})\}$, $k = 0, ..., n$. Since V_t is unknown and H_n^0 cannot be calculated, we approximate V_t using observational data. This model is a type of backward SDE.

- (A general expression of backward SDE):

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)\, ds - \int_t^T Z_s\, dW_s,$$

where ξ is a known condition (terminal value), and Z_t is an unknown process.

Applications of BSDEs

- (Optimal asset allocation)
  When we model stock price processes with SDEs and consider some specific utility function, the optimal wealth process Y_t is given by

  $$dY_t = (-F(X_t, V_t) + \lambda)dt + V_t dW_t$$

  for some X_t, V_t, F and λ.
- Then, optimal strategy is associated with the function F of BSDE. (Chong et al. (2019))
- Ergodic BSDEs also have applications in the field of stochastic control problem. (Richou (2019))

Proposed Estimator

We construct the maximum likelihood-type estimator $\hat{\theta}_n$ as follows: Let $(c_n)_{n=1}^{\infty}$ be a positive integer sequence. Define $L_n = [n/c_n], t_m^l = (m + c_n l)h_n$, and estimate $Z_t = V_t V_t^\top$ as follows:

$$\hat{Z}_l = \frac{1}{c_n h_n} \sum_{m=1}^{c_n} (Y_{t_m^l} - Y_{t_{m-1}^l})(Y_{t_m^l} - Y_{t_{m-1}^l})^\top, \quad (0 \leq l \leq L_n - 1).$$

- Using c_n observations to create an estimator for $Z_{t_0^l}$

Proposed Estimator

Then, define the quasi-log-likelihood $H_n(\theta)$ as

$$H_n(\theta) = -\frac{1}{2}\sum_{l=1}^{L_n-1}\sum_{m=1}^{c_n}\left(\bar{Y}_m^l\right)^\top(\hat{Z}_{l-1}h_n)^{-1}\bar{Y}_m^l 1_{\{\det \hat{Z}_{l-1}>0\}}$$

where $\bar{Y}_m^l = Y_{t_m^l} - Y_{t_{m-1}^l} - h_n\hat{\psi}_{l,m}(\theta)$,
$\hat{\psi}_{l,m}(\theta) = \psi(X_{t_{m-1}^l}, Y_{t_{m-1}^l}, \hat{Z}_{l-1}, \theta)$.

Proposed Estimator

Then, define the quasi-log-likelihood $H_n(\theta)$ as

$$H_n(\theta) = -\frac{1}{2} \sum_{l=1}^{L_n-1} \sum_{m=1}^{c_n} \left(\bar{Y}_m^l\right)^\top (\hat{Z}_{l-1} h_n)^{-1} \bar{Y}_m^l 1_{\{\det \hat{Z}_{l-1} > 0\}}$$

where $\bar{Y}_m^l = Y_{t_m^l} - Y_{t_{m-1}^l} - h_n \hat{\psi}_{l,m}(\theta)$,
$\hat{\psi}_{l,m}(\theta) = \psi(X_{t_{m-1}^l}, Y_{t_{m-1}^l}, \hat{Z}_{l-1}, \theta)$.

- $\hat{Z}_{l-1}$ corresponds to $b_k b_k^\top$ in H_n^0.
- $(c_n)_{n=1}^\infty$ should satisfy

$$c_n n^{-\epsilon} \to \infty, \quad c_n h_n n^\epsilon \to 0, \quad n c_n^2 h_n^3 \to 0, \quad \frac{\sqrt{n h_n}}{c_n} \to 0 \qquad (2)$$

for some $\epsilon > 0$. (For example, $c_n = 1 + [h_n^{-1/2}]$)

Ergodicity

Define the maximum likelihood-type estimator as: $\hat{\theta}_n \in \mathrm{argmax}_\theta H_n(\theta)$. Assume ergodicity.

1. When $\psi(x, y, z, \theta)$ does not depend on y: There exists a probability distribution $\pi(x, z)$ such that for any π-integrable function f,

$$\frac{1}{T}\int_0^T f(X_t, V_t V_t^\top)dt \xrightarrow{P} \int f(x, z)\pi(dxdz), \quad (T \to \infty).$$

Ergodicity

Define the maximum likelihood-type estimator as: $\hat{\theta}_n \in \mathrm{argmax}_\theta H_n(\theta)$.
Assume ergodicity.

1. When $\psi(x, y, z, \theta)$ does not depend on y: There exists a probability distribution $\pi(x, z)$ such that for any π-integrable function f,

$$\frac{1}{T}\int_0^T f(X_t, V_t V_t^\top)dt \xrightarrow{P} \int f(x, z)\pi(dxdz), \quad (T \to \infty).$$

2. Otherwise: There exists a probability distribution $\pi(x, y, z)$ such that for any π-integrable function f,

$$\frac{1}{T}\int_0^T f(X_t, Y_t, V_t V_t^\top)dt \xrightarrow{P} \int f(x, y, z)\pi(dxdydz), \quad (T \to \infty).$$

Main Theorem

Define

$$\hat{\theta}_n \in \mathrm{argmax}_\theta H_n(\theta),$$

$$\Gamma = \int \partial_\theta \psi(x, y, z, \theta_0)^\top z^{-1} \partial_\theta \psi(x, y, z, \theta_0) d\pi$$

Theorem 2 (Asymptotic Normality)

Under ergodicity and conditions on smoothness, non-degeneracy of a, b, and moment conditions of X_t, V_t, etc.,

$$\sqrt{nh_n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Gamma^{-1}), \quad (n \to \infty).$$

Remark

- While Masuda (2005) proposed a least-square-type estimator for drift term estimation when V_t is unknown, the advantage of the proposed estimator is that it can estimate even when the unknown V_t is included in the drift term ψ.

Remark

- While Masuda (2005) proposed a least-square-type estimator for drift term estimation when V_t is unknown, the advantage of the proposed estimator is that it can estimate even when the unknown V_t is included in the drift term ψ.
- If we consider an auxiliary model observing $(Z_{kh_n})_{k=0}^n$ in addition to $(X_{kh_n}, Y_{kh_n})_{k=0}^n$, it is included in the settings of previous studies by Kessler (1997) and Yoshida (2011).
  - Gobet (2002) showed the local asymptotic normality of the auxiliary model and proved that the optimal asymptotic variance of the estimator is Γ^{-1}.
  - In other words, the asymptotic variance of $\hat{\theta}_n$ is optimal in the sense that it achieves the lower bound in the model with additional observations.

Key Points of the Proof

- Unlike existing studies, even if we assume $\inf_t \det(V_t {V_t}^\top) > 0$, the approximation $\hat{Z}_l$ of the volatility $V_t {V_t}^\top$ is not guaranteed to satisfy $\det \hat{Z}_l > 0$ and may degenerate.
    - We handle this using control with stopping times and martingale evaluations for sums with stopping times.

Key Points of the Proof

- Unlike existing studies, even if we assume $\inf_t \det(V_t V_t^\top) > 0$, the approximation $\hat{Z}_l$ of the volatility $V_t V_t^\top$ is not guaranteed to satisfy $\det \hat{Z}_l > 0$ and may degenerate.
  - We handle this using control with stopping times and martingale evaluations for sums with stopping times.

- In the calculation of $\hat{Z}_l$, as c_n increases, the approximation error of $\int_{t_0^l}^{t_0^{l+1}} V_t V_t^\top dt$ decreases, but the error from using $\hat{Z}_{l-1}$ instead of $\hat{Z}_l$ increases.
  - Appropriate settings considering the trade-off of c_n are necessary, making it much more difficult than usual Euler approximation.
  - As a result, we obtain the same optimal variance as before, and can prove it under the same condition $nh_n^2 \to 0$ for h_n.

Example (Stochastic Volatility Model)

Consider a stochastic process Y_t satisfying the following:

$$dY_t = \psi(t, Y_t, \theta)dt + V_t dW_t.$$

Here, V_t is an unknown stochastic process, and (X_t, V_t) satisfies ergodicity.

Example (Stochastic Volatility Model)

Consider a stochastic process Y_t satisfying the following:

$$dY_t = \psi(t, Y_t, \theta)dt + V_t dW_t.$$

Here, V_t is an unknown stochastic process, and (X_t, V_t) satisfies ergodicity.

While there was no theory for maximum likelihood-type estimators of θ_0 in such cases where V_t has no assumed parametric model, our proposed estimator provides asymptotic normality and other theoretical results.

Example (Stochastic Volatility Model)

Consider a stochastic process Y_t satisfying the following:

$$dY_t = \psi(t, Y_t, \theta)dt + V_t dW_t.$$

Here, V_t is an unknown stochastic process, and (X_t, V_t) satisfies ergodicity.

While there was no theory for maximum likelihood-type estimators of θ_0 in such cases where V_t has no assumed parametric model, our proposed estimator provides asymptotic normality and other theoretical results.

As this model includes stochastic volatility models commonly used for stock prices, it enables the estimation of drift terms in stochastic volatility.

Example (Chong et al. (2019))

Chong et al. (2019) consider the following BSDE model to maximize the utility function of a portfolio:

$$dY_t = (F(X_t, V_t, \gamma) - \lambda)dt + V_t dW_t.$$

Under assumptions such as ergodicity of (X_t, V_t) and smoothness of F, our proposed estimator enables the estimation of parameters γ and λ.

Numerical Experiment

We conduct numerical experiments with the following model:

$$dY_t = \theta_0 \sqrt{X_t^2 + 0.1} dt + \sqrt{X_t^2 + 0.1} dW_t,$$
$$dX_t = a(b - X_t)dt + \sigma dW_t.$$

The parameters are set as $a = 2$, $b = 0.3$, $\sigma = 0.025, \theta_0 = 10$, with initial values $X_0 = 0.3$, $Y_0 = 1$.

Numerical Experiment

We conduct numerical experiments with the following model:

$$\begin{aligned} dY_t &= \theta_0 \sqrt{X_t^2 + 0.1} dt + \sqrt{X_t^2 + 0.1} dW_t, \\ dX_t &= a(b - X_t)dt + \sigma dW_t. \end{aligned}$$

The parameters are set as $a = 2$, $b = 0.3$, $\sigma = 0.025, \theta_0 = 10$, with initial values $X_0 = 0.3$, $Y_0 = 1$.

We run 100 simulations and calculate the ML-type estimator, and its error at $n = 100,000$:

$$\text{Error} = \frac{|\hat{\theta}_n - \theta_0|}{\theta_0}.$$

Numerical Experiment

We define the observation width h_n and the number of observations c_n for calculating $\hat{Z}_l$ for integers k, l as follows:

$$c_n = n^{0.05k}, \quad k = 1, 2, \ldots, l-1$$

$$h_n = n^{-0.05l}, \quad l = 1, 2, \cdots$$

Numerical Experiment

We define the observation width h_n and the number of observations c_n for calculating $\hat{Z}_l$ for integers k, l as follows:

$$c_n = n^{0.05k}, \quad k = 1, 2, \ldots, l-1$$

$$h_n = n^{-0.05l}, \quad l = 1, 2, \cdots$$

To satisfy the conditions (2) for c_n, h_n, l and k must satisfy:

$$11 \leq l \leq 19, \quad 10 - \frac{l}{2} < k < 1.5l - 10.$$

($nc_n^2 h_n^3 \to 0 \Rightarrow k < 1.5l - 10, \quad \frac{\sqrt{nh_n}}{c_n} \to 0 \Rightarrow 10 - \frac{l}{2} < k,$
$nh_n^2 \to 0 \Rightarrow 10 < l$)

Numerical Experiment

		l				
		11	12	13	14	15
k	3					0.9295
	4			0.3797	0.4404	0.4214
	5	0.1308	0.1739	0.2274	0.2536	0.3065
	6	0.0604	0.0868	0.1241	0.1743	0.2099
	7		0.0810	0.1095	0.1630	0.1954
	8		0.0765	0.1003	0.1281	0.1978
	9			0.1165	0.1389	0.1978
	10				0.1367	0.1753
	11				0.1323	0.2178
	12					0.2296

Table: The values of error for each k, l

Numerical Experiment

- We can confirm that the estimation error is kept small for many combinations of k, l. The convergence rate $(nh_n)^{-1/2}$ derived in Theorem 2 is not very fast, being $(nh_n)^{-1/2} = n^{-0.225}$ even for the best case of $l = 11$.

Numerical Experiment

- We can confirm that the estimation error is kept small for many combinations of k, l. The convergence rate $(nh_n)^{-1/2}$ derived in Theorem 2 is not very fast, being $(nh_n)^{-1/2} = n^{-0.225}$ even for the best case of $l = 11$.

- The choice of l is shown to have a significant impact on the estimation accuracy, which is consistent with the convergence rate $(nh_n)^{-1/2}$. The pair $(l, k) = (11, 6)$ gives the minimum error, providing the most accurate estimation of θ.

Numerical Experiment

- We can confirm that the estimation error is kept small for many combinations of k, l. The convergence rate $(nh_n)^{-1/2}$ derived in Theorem 2 is not very fast, being $(nh_n)^{-1/2} = n^{-0.225}$ even for the best case of $l = 11$.

- The choice of l is shown to have a significant impact on the estimation accuracy, which is consistent with the convergence rate $(nh_n)^{-1/2}$. The pair $(l, k) = (11, 6)$ gives the minimum error, providing the most accurate estimation of θ.

- For $l = 16$, $k = 13$ is best with ERROR$=0.2298$,
For $l = 17$, $k = 8$ is best with ERROR$=0.3160$,
For $l = 18$, $k = 13$ is best with ERROR$=0.3892$,
For $l = 19$, $k = 14$ is best with ERROR$=0.5232$.

Summary

- We constructed a maximum likelihood-type estimator for the drift parameter θ_0 in a setting including backward SDEs, where the volatility term V_t is unobserved and no parametric model is assumed.
- Under assumptions such as ergodicity, we demonstrated the consistency and asymptotic normality of the estimator, and confirmed that it achieves the optimal asymptotic variance as in the case where V_t is observed.
- In numerical experiments, we confirmed that the estimation error decreases for large sample sizes. While the theoretically optimal rate for the number of observations c_n used to estimate V_t is unknown, numerical experiments suggest that a rate slightly larger than $h_n^{-1/2}$ tends to be best.

References

1. Chong, W. F., Hu, Y., Liang, G., & Zariphopoulou, T. (2019). An ergodic BSDE approach to forward entropic risk measures: Representation and large-maturity behavior. Finance and Stochastics, 23(1), 239-273.
2. Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, 24(2), 211-229.
3. Masuda, H. (2005). Simple estimators for parametric Markovian trend of ergodic processes based on sampled data. Journal of the Japan Statistical Society, 35(2), 147-170.
4. Ogihara, T. and Stadje, M. (2024). Efficient drift parameter estimation for ergodic solutions of backward SDEs, Scandinavian Journal of Statistics, 51(3), 1181–1205.
5. Richou, A. (2009). Ergodic BSDEs and related PDEs with Neumann boundary conditions. Stochastic Processes and their Applications, 119(9), 2945–2969.
6. Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. Annals of the Institute of Statistical Mathematics, 63(3), 431-479.