Journal of Econometrics 191 (2016) 110–128

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Information theory for maximum likelihood estimation of diffusion models[☆]

Hwan-sik Choi

Department of Economics, Binghamton University, 4400 Vestal Parkway E. P.O. Box 6000, Binghamton, NY 13902, USA

ARTICLE INFO

Article history:
Received 7 March 2014
Received in revised form
8 March 2015
Accepted 5 October 2015
Available online 23 October 2015

JEL classification:
C13
C22
C40

Keywords:
Kullback–Leibler information criterion
Fiber bundle
High frequency
Infinitesimal generator
Infill asymptotics
Misspecification

ABSTRACT

We develop an information theoretic framework for maximum likelihood estimation of diffusion models. Two information criteria that measure the divergence of a diffusion process from the true diffusion are defined. The maximum likelihood estimator (MLE) converges asymptotically to the limit that minimizes the criteria when sampling interval decreases as sampling span increases. When both drift and diffusion specifications are correct, the criteria become zero and the inverse curvatures of the criteria equal the asymptotic variance of the MLE. For misspecified models, the minimizer of the criteria defines pseudo-true parameters. Pseudo-true drift parameters depend on approximate transition densities if used.

Published by Elsevier B.V.

1. Introduction

We develop an information theoretic framework for maximum likelihood estimation of diffusion models in an asymptotic environment in which sampling interval shrinks to zero as sampling span increases to infinity. The new approach provides a unified framework within which we can analyze both correctly specified and misspecified diffusion models.

For discrete time models, White (1982) gives an information theoretic foundation for analyzing the maximum likelihood estimator (MLE). His theory is suitable for the models analyzed with the conventional large sample theory with fixed sampling interval. For diffusion models, estimation and inference under decreasing sampling interval were studied previously by, for example, Bandi and Phillips (2003), Tang and Chen (2009), Chang and Chen (2011), Jeong and Park (2013), and Choi et al. (2014). Specifically, Jeong and Park (2013) study the maximum likelihood estimation for correctly specified parametric models for both stationary and non-stationary processes. Choi et al. (2014) provide some asymptotic results for misspecified models, although their main focus is on model selection testing.

The theory developed in Jeong and Park (2013) and Choi et al. (2014) can explain the differential properties of the drift and diffusion parameter estimators observed in high frequency samples. They show that the rates of information accumulation for drift and diffusion parameters are different, and both sample size and sampling span are important for diffusion parameter estimation whereas only sampling span matters for drift parameter estimation. The present paper provides an information theoretic foundation for their new asymptotic theory.

In White's theory, the Kullback–Leibler information criterion (KLIC, Kullback and Leibler (1951)) plays a key role in understanding the asymptotic properties of the MLE. He shows that the MLE converges to the limit that minimizes the KLIC from the true distribution and derives asymptotic properties of the MLE under model misspecification. He also shows that the usual inference procedure with the Likelihood Ratio, Lagrange Multiplier, and Wald test

☆ I thank Minsoo Jeong, Joon Park, Jose Figueroa-Lopez, James MacKinnon, Nick Kiefer, Yacine Aït-Sahalia, an Associate Editor, an anonymous referee, and the seminar participants at Purdue (Statistics), Cornell, Queen's University, Washington University in St. Louis (Mathematics), NY Camp Econometrics, Symposium of Probability and Stochastic Processes (CIMAT, Mexico), the Infometrics Conference (American University), and the 2015 Winter Meeting of the Econometric Society for helpful discussions.

E-mail address: hwansik@binghamton.edu.

http://dx.doi.org/10.1016/j.jeconom.2015.10.002
0304-4076/Published by Elsevier B.V.

statistics is not valid for a misspecified model. Naturally, we are interested in the questions similar to those that White has addressed in his paper. Under the new asymptotic framework with decreasing sampling interval, do the MLEs of misspecified diffusion models converge to some limits asymptotically, do the limits have any meaning, are the usual inference procedures based on likelihood ratios valid, what is the consequence of misspecification in either the drift or the diffusion function only, what is the impact of having to use an approximate transition density when the exact transition density is unknown in closed-form? We provide an answer to each of these questions in this paper.

Based on the KLIC, we develop two-tier information criteria, which we refer to as the primary and secondary information criteria. The criteria measure the divergence of a diffusion process M from the true diffusion X. The primary criterion depends on the diffusion function of M, and equals zero if and only if the diffusion function of M is the same as that of X. The secondary criterion becomes zero if both drift and diffusion functions of M are correct.

Our new information criteria play a similar role as the KLIC in White (1982). When sampling interval decreases as sampling span increases, the MLE asymptotically converges to the limit that minimizes the two criteria. For correctly specified diffusion models, the two criteria are minimized at the true parameter value, therefore, the MLE is consistent. However, for misspecified diffusion models, there is no parameter value that makes both criteria exactly zero. In this case, the minimizer of the criteria is said to be the pseudo-true parameter and the MLE converges to this parameter asymptotically.

Because the primary criterion depends on diffusion specification, but not on drift specification of a model, the primary criterion can be minimized to zero as long as diffusion function is correctly specified. Consequently, the diffusion function estimator is consistent regardless of drift function misspecification. However, the secondary criterion depends on both diffusion and drift specifications. Therefore, even if drift function specification is correct, the MLE of the drift parameter may not be consistent when diffusion function is misspecified.

In most applications, we do not know the exact transition density of a model in closed-form. Hence, we must use an approximate transition density. We show that the primary criterion is invariant to approximate transition densities that satisfy some regularity conditions that are always met by the exact transition density. Especially, the primary criterion is the same for the three popular approximate transition densities, the Euler, Milstein, and Aït-Sahalia's closed-form approximate transition densities. However, the secondary criterion depends on the choice of an approximate transition density. This implies that the pseudo-true drift parameter depends on the choice of an approximate transition density if used. An exception to this occurs when diffusion function is correctly specified. Under correct diffusion function specification, the secondary criterion has the same form for the exact and the popular three approximate transition densities mentioned above.

As explained previously, when drift function is correctly specified but diffusion function is misspecified, the drift estimator is not consistent in general. But we show that the Euler approximate transition density has a robustness property in the sense that the MLE of the drift parameter converges asymptotically to the true value regardless of the degree of misspecification in the diffusion function. For the Milstein and Aït-Sahalia's closed-form approximate transition densities, the MLE of the drift parameter may not converge to the true parameter value. A Monte Carlo study is given in Section 3.3 to demonstrate this phenomenon.

We also study the asymptotic distribution of the MLE for both correctly specified and misspecified models. When a diffusion model is correctly specified, we show that the asymptotic variances of the MLE of diffusion and drift parameters are given by the inverse curvatures of the primary and secondary criteria, respectively. For misspecified models, the asymptotic variance of the MLE depends on both expected Hessian matrix and covariance matrix of score functions. Based on this result, a misspecification test using the information matrix identity is developed in Section 3.2.

The theory developed in this paper has many potential applications in finance and economics. They include estimation and specification testing of interest rates (Aït-Sahalia, 1999), volatility (Aït-Sahalia and Kimmel, 2007), and term structure models (Aït-Sahalia and Kimmel, 2010). Our theory is applicable to partially or completely misspecified models as well as correctly specified models, and therefore, it provides empirical researchers with a more general perspective and new insights on the maximum likelihood estimation and inference of the diffusion models widely used in various financial and economic applications.

We develop the new information criteria in the next section, and analyze the asymptotic properties of the MLE in an information theoretic framework in Section 3.

2. Divergence of diffusion processes

Let W be the standard Brownian motion defined on a probability space $(\Omega, \mathfrak{F}, \mathbf{P})$, and X be a stationary diffusion on $\mathcal{D} \subset \mathbb{R}$ that solves the stochastic differential equation (SDE)

$$dX_t = \mu_0(X_t)\,dt + \sigma_0(X_t)\,dW_t \tag{2.1}$$

with a drift function $\mu_0(\cdot)$ and a diffusion function $\sigma_0(\cdot)$. We assume that SDEs in this paper satisfy the conditions in Karatzas and Shreve (1991) to admit a weak solution. Suppose we have a parametric diffusion model $M(\theta)$ for X given by

$$M(\theta) : dX_t = \mu(X_t; \alpha)\,dt + \sigma(X_t; \beta)\,dW_t, \tag{2.2}$$

where $\mu(\cdot; \alpha)$ and $\sigma(\cdot; \beta) > 0$ are known functions with an unknown parameter vector $\theta = (\alpha, \beta)$ in a compact set $\Theta \subset \mathbb{R}^k$. A diffusion model $M(\theta)$ is misspecified if there exists no $\theta \in \Theta$ such that $\mu_0(\cdot) = \mu(\cdot; \alpha)$ and $\sigma_0(\cdot) = \sigma(\cdot; \beta)$ on $\mathcal{D}$. In this paper, we separate the drift parameter α from the diffusion parameter β because their asymptotic properties are quite different. If a parameter appears in both $\mu(\cdot)$ and $\sigma(\cdot)$ as in transformed diffusion models such as Bu et al. (2011) and Forman and Sørensen (2014), the theory developed in this paper suggests that the asymptotic theory for β would apply to such parameter.

Let $p_0(t, x, y)$ be the transition density from $X_0 = x$ to $X_t = y$ of the true process X that solves (2.1) and $p(t, x, y; \theta)$ be the exact or an approximate transition density of $M(\theta)$. To define how faraway $M(\theta)$ is from X, we consider the KLIC from the transition density $p_0(t, x, y)$ to $p(t, x, y; \theta)$. The KLIC(P, Q) from a probability measure P to an equivalent (mutually absolutely continuous) probability measure Q, is defined by

$$\text{KLIC}(P, Q) \equiv \int \log\left(\frac{dP}{dQ}\right) dP,$$

and it is infinite if P and Q are not equivalent. See Csiszár (1967a,b, 1975) for more general divergence measures.

Based on the KLIC, define the divergence measure

$$D_t(\theta) \equiv \mathbf{E}\left(\log \frac{p_0(t, X_0, X_t)}{p(t, X_0, X_t; \theta)}\right), \tag{2.3}$$

where the expectation is taken with respect to the true diffusion unconditionally. The function $D_t(\theta)$ characterizes the divergence from the true process X to $M(\theta)$ at various transition intervals. For instance, $D_t(\theta)$ for small t would measure how close the model transition density $p(t, x, y; \theta)$ is from the true transition density $p_0(t, x, y)$ for observations X_0 and X_t obtained over a small time interval t. In the limit of $t \to \infty$, $D_\infty(\theta)$ would measure the divergence of the marginal densities of X and $M(\theta)$.