Identification problem of the exponential tilting estimator under misspecification*

Naoya Sueishi

Graduate School of Economics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan.
Phone:+81-75-753-3500. Fax:+81-75-753-3492.

Abstract

This paper shows a potential pitfall of the exponential tilting (ET) estimator under misspecification. We show that the pseudo-true value of the ET estimator is not identified if the true distribution is not absolutely continuous with respect to the probability measures implied by the moment restriction model. This result implies that the ET estimator cannot be consistent for the pseudo-true value if the moment generating function of the moment function is unbounded.

Keywords: Exponential tilting estimator, Misspecified models, *I*-projection.
JEL classification: C13, C14

1. Introduction

There have been numerous studies on inference of misspecified parametric models. This is partially because of viewpoint that most econometric models are only approximations to reality and thus intrinsically misspecified. The use of the Kullback-Leibler information criterion (KLIC) enables us to make a reasonable inference even for misspecified models. For instance, White (1982) shows that the maximum likelihood estimator is $\sqrt{n}$-consistent for

the pseudo-true value, which minimizes the KLIC from the model to the true distribution. Vuong (1989) proposes a testing procedure for comparing two potentially misspecified models. Sin and White (1996) discuss the properties of the information criteria for selecting misspecified models.

Recently, there has also been a growing interest in inference of misspecified moment restriction models. Hall and Inoue (2003) have studied the large sample properties of the generalized method of moments estimator (Hansen (1982)) under misspecification. Kitamura (2001) provides a Vuong-type test by using the exponential tilting (ET) estimator (Kitamura and Stutzer (1997), Imbens et al. (1998)). Chen et al. (2007) propose a nonparametric likelihood ratio test for choosing between parametric and moment restriction models on the basis of the empirical likelihood (EL) estimator (Qin and Lawless (1994)). The EL and ET estimators are well-suited for analyzing misspecified moment restriction models because their pseudo-true values are naturally defined through the KLIC minimization problem.

This paper shows a potential pitfall of the ET estimator under misspecification. Schennach (2007) shows that the EL estimator ceases to be $\sqrt{n}$-consistent for any parameter value if the moment function is unbounded. The proof of the theorem suggests that the ET estimator suffers from the same problem if the moment generating function (MGF) of the moment function is unbounded. Although Schennach (2007) claims that her theorem does not prevent the EL estimator from being a consistent estimator for the pseudo-true value, she leaves an open question: what is the pseudo-true value? We show that the pseudo-true value of the ET estimator is not identified if the MGF is unbounded. Thus, the ET estimator cannot even be consistent for the pseudo-true value. The main problem is that if the model is misspecified, then the set of probability measures implied by the model may not have a common support with the true distribution. If this is the case, then the ET estimator no longer solves the sample counterpart of the KLIC minimization problem.

2. Duality theorem and ET estimator

Let $x_1, \ldots, x_n$ be an i.i.d. sequence of random vectors from an unknown distribution μ with support $\mathcal{X} \subset \mathbb{R}^{d_x}$. The model we consider takes the form

$$\mathrm{E}_\mu[g(x_i, \theta_0)] = \int g(x, \theta_0)d\mu = 0, \tag{1}$$

2

where $g : \mathcal{X} \times \Theta \to \mathbb{R}^l$ is a known function up to the parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$. We assume $l > p$.

Model (1) implies a nonparametric set of probability measures. Let $\mathbf{M}$ be the set of all probability measures on $\mathbb{R}^{d_x}$, and let $\mathcal{P}_\theta = \{P \in \mathbf{M} : \int g(x,\theta) dP = 0\}$. Then, we define $\mathcal{P} = \cup_{\theta \in \Theta} \mathcal{P}_\theta$, which is the set of all probability measures compatible with moment restriction. We call $\mathcal{P}$ the moment restriction model. The model $\mathcal{P}$ is correctly specified if $\mu \in \mathcal{P}$, and misspecified otherwise.

The KLIC or I-divergence from $\mathcal{P}$ to μ is defined as $D(\mathcal{P}\|\mu) = \inf_{P \in \mathcal{P}} D(P\|\mu)$, where

$$D(P\|\mu) = \begin{cases} \int \log\left(\frac{dP}{d\mu}\right) dP & \text{if } P \ll \mu \\ \infty & \text{otherwise.} \end{cases}$$

The KLIC defines a pseudo-distance between the model and the true distribution. If there exists $P^* \in \mathcal{P}$ such that $D(P^*\|\mu) = D(\mathcal{P}\|\mu)$, then P^* is called the I-projection; see, for example, Csiszár (1975). The I-projection is the closest measure to μ among $\mathcal{P}$.

The KLIC is typically obtained in two steps. First, for fixed θ, we solve

$$D(\mathcal{P}_\theta\|\mu) = \inf_P \int \log\left(\frac{dP}{d\mu}\right) dP \quad \text{s.t.} \int g(x,\theta) dP = 0 \text{ and } \int dP = 1. \quad (2)$$

The primal problem (2) has the following dual problem:

$$D^*(\mathcal{P}_\theta\|\mu) = \max_{\lambda \in \mathbb{R}^l} \left[-\log\left(\int \exp\left(\lambda' g(x,\theta)\right) d\mu \right) \right]. \quad (3)$$

The Fenchel duality theorem (see, e.g., Borwein and Lewis (1991) and Kitamura (2007)) implies that $D(\mathcal{P}_\theta\|\mu) = D^*(\mathcal{P}_\theta\|\mu)$. Hence, we get

$$D(\mathcal{P}\|\mu) = \inf_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^l} \left[-\log\left(\int \exp(\lambda' g(x,\theta)) d\mu \right) \right]. \quad (4)$$

We define the pseudo-true value (θ^*, λ^*) as the value of the parameter that attains the minimum of KLIC:

$$(\theta^*, \lambda^*) = \arg\max_{\theta \in \Theta} \arg\min_{\lambda \in \mathbb{R}^l} \left[\int \exp\left(\lambda' g(x,\theta)\right) d\mu \right]. \quad (5)$$

The μ-density of the I-projection is obtained by

$$\frac{dP^*}{d\mu} = \frac{\exp(\lambda^{*\prime} g(x, \theta^*))}{\int \exp(\lambda^{*\prime} g(x, \theta^*)) d\mu}. \tag{6}$$

The ET estimator $(\hat{\theta}, \hat{\lambda})$ solves the sample analogue of (5):

$$\max_{\theta \in \Theta} \min_{\lambda \in \mathbb{R}^l} \left[\frac{1}{n} \sum_{i=1}^{n} \exp\left(\lambda' g(x_i, \theta)\right) \right]. \tag{7}$$

Christoffersen et al. (2001) show that the ET estimator is $\sqrt{n}$-consistent for the pseudo-true value under certain regularity conditions.

Schennach (2007) shows that the EL estimator is not $\sqrt{n}$-consistent for any value of $\theta \in \Theta$ if $\|g(x, \theta)\|$ is unbounded for all $\theta \in \Theta$. The key problem is described as follows: The population dual problem of the EL estimator is given by $\min_{\theta \in \Theta} \max_{\lambda \in \mathbb{R}^l} \mathrm{E}_\mu[\log(1 - \lambda' g(x_i, \theta))]$. However, if $\|g(x, \theta)\|$ is unbounded, then $\mathrm{E}_\mu[\log(1 - \lambda' g(x_i, \theta))]$ is ill-defined for $\lambda \neq 0$, which makes the behavior of the EL estimator unusual. The proof of her theorem also suggests that the ET estimator suffers from the same problem if the MGF of the moment function is unbounded, that is, $\mathrm{E}_\mu[\exp(\lambda' g(x_i, \theta))]$ is unbounded for all $\theta \in \Theta$ and $\lambda \neq 0$. If $\mathrm{E}_\mu[\exp(\lambda' g(x_i, \theta))]$ is unbounded, then the only possible value of λ is zero, which forces $\hat{\lambda} \xrightarrow{p} 0$. Then, we can show that the ET estimator cannot be $\sqrt{n}$-consistent for any $\theta \in \Theta$ in a similar fashion as in the proof of Schennach (2007).

We explain this phenomenon from an identification point of view. Schennach (2007) proves the result under the assumption that the pseudo-true value exists. On the other hand, we show that the pseudo-true value of the ET estimator is not identified. The above argument suggests that zero is a candidate for the pseudo-true value of λ. However, $\lambda^* = 0$ and (6) imply that the model is correctly specified ($P^* = \mu$). Thus, λ^* must be non-zero. As we will see in the next section, the problem resides in the fact that $\mathcal{P}_\theta$ and μ do not have a common support. Then, the equivalence between (2) and (3) does not hold anymore. The ET estimator no longer solves the sample counterpart of the KLIC minimization problem. Thus, the ET estimator is not even consistent for the pseudo-true value.

3. Main result

The existence of the I-projection is a subtle problem. A sufficient condition for the existence is that $\mathcal{P}$ is variation-closed (Csiszár (1975)). For

instance, if $\sup_{\theta \in \Theta, x \in \mathcal{X}} \|g(x, \theta)\| < \infty$, then the I-projection exists. Even when the I-projection does not exist, the generalized I-projection (Csiszár (1984)) exists under a fairly weak assumption. The generalized I-projection is defined as follows:

Definition 3.1. *If each sequence $\{P_j\}$ in $\mathcal{P}$ satisfying $D(P_j \| \mu) \to D(\mathcal{P} \| \mu)$ converges P^* in variation, then P^* is called the generalized I-projection.*

Intuitively speaking, the generalized I-projection is the closest measure to μ in the closure of $\mathcal{P}$. If $P^* \in \mathcal{P}$, then it coincides with the I-projection.

Before stating the main result, we introduce the following notations. For fixed θ, let $\mathcal{C}_\theta^k = \left\{ x \in \mathbb{R}^{d_x} : \|g(x, \theta)\| \leq k \right\}$. We define

$$\mathcal{P}_\theta^0 = \bigcup_{k \in \mathbb{N}} \left\{ P \in \mathbf{M} : \int_{\mathcal{C}_\theta^k} g(x, \theta) dP = 0 \text{ and } P(\mathcal{C}_\theta^k) = 1 \right\},$$

which is a subset of $\mathcal{P}_\theta$. Note that $\|g(x, \theta)\|$ is bounded on the support of $P \in \mathcal{P}_\theta^0$. In addition, let $\mathcal{C}_\theta$ be the smallest linear subspace of $\mathbb{R}^{d_x}$ with the property that $P(\mathcal{C}_\theta) = 1$ for all $P \in \mathcal{P}_\theta$ with $P \ll \mu$. Let P_θ^* denote the generalized I-projection of μ onto $\mathcal{P}_\theta$. The following theorem is a simple application of Csiszár (1984).

Theorem 3.1. *If there exists $P \in \mathcal{P}_\theta$ such that $P \ll \mu$, then $D(\mathcal{P}_\theta \| \mu) = D(\mathcal{P}_\theta^0 \| \mu) < \infty$. Also, P_θ^* exists and has the μ-density of the form*

$$\frac{dP_\theta^*}{d\mu} = \begin{cases} \exp\left\{ D(\mathcal{P}_\theta \| \mu) + \gamma' g(x, \theta) \right\} & \text{if } x \in \mathcal{C}_\theta \\ 0 & \text{otherwise} \end{cases}$$

for some $\gamma \in \mathbb{R}^l$. Moreover, we obtain

$$D(\mathcal{P}_\theta \| \mu) = \max_{\lambda \in \mathbb{R}^l} \left[-\log \left(\int \exp\left(\lambda' g(x, \theta) \right) d\mu \right) \right]$$

if and only if there exists $P \in \mathcal{P}_\theta$ such that P is mutually absolutely continuous with respect to μ.

The proof is omitted since it is clearly understood from Csiszár (1984).

The theorem states that if there is no $P \in \mathcal{P}$ such that P is mutually absolutely continuous with respect to μ, then (4) does not hold. Although the theorem does not necessarily prevent the existence of the pseudo-true

value of the ET estimator, the pseudo-true value is not identified by (5). The saddle point problem (7) is not the dual problem for the sample counterpart of the KLIC minimization problem. This problem will occur if the moment restriction is satisfied only by the probability measures under which the moment function is bounded, while the moment function is unbounded under the true distribution. The condition that the MGF is unbounded is a sufficient condition for this phenomenon. We conjecture that a similar result also holds for the EL estimator.

References

Borwein, J.M., Lewis, A.S., 1991. Duality relationships for entropy-like minimization problems. SIAM Journal on Control and Optimization 29, 325–338.

Chen, X., Hong, H., Shum, M., 2007. Nonparametric likelihood ratio model selection tests between parametric likelihood and moment condition models. Journal of Econometrics 141, 109–140.

Christoffersen, P., Hahn, J., Inoue, A., 2001. Testing and comparing value-at-risk measures. Journal of Empirical Finance 8, 325–342.

Csiszár, I., 1975. i-divergence geometry of probability distributions and minimization problems. Annals of Probability 3, 146–158.

Csiszár, I., 1984. Sanov property, generalized i-projection and a conditional limit theorem. Annals of Probability 12, 768–793.

Hall, A.R., Inoue, A., 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. Journal of Econometrics 114, 361–394.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Imbens, G.W., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Kitamura, Y., 2001. Comparing misspecified dynamic econometric models using nonparametric likelihood. Unpublished Manuscript, University of Wisconsin.

Kitamura, Y., 2007. Empirical likelihood methods in econometrics: Theory and practice, in: Blundell, R., Newey, W.K., Persson, T. (Eds.), Advances in Economics and Econometrics. Cambridge University Press, pp. 174–237.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Schennach, S.M., 2007. Point estimation with exponentially tilted empirical likelihood. Annals of Statistics 35, 634–672.

Sin, C., White, H., 1996. Information criteria for selecting possibly misspecified parametric models. Journal of Econometrics 71, 207–225.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

White, H., 1982. Maximum likelihood estimation for misspecified models. Econometrica 50, 1–25.