A Simple Derivation of the Efficiency Bound for Conditional Moment Restriction Models

Naoya Sueishi

September 2015
Discussion Paper No.1531

GRADUATE SCHOOL OF ECONOMICS
KOBE UNIVERSITY

ROKKO, KOBE, JAPAN

A Simple Derivation of the Efficiency Bound for Conditional Moment Restriction Models*

Naoya Sueishi
Kobe University†

September 2015

Abstract

This study gives a simple derivation of the efficiency bound for conditional moment restriction models. The Fisher information is obtained by deriving a least favorable submodel in an explicit form. The proposed method also suggests an asymptotically efficient estimator, which can be viewed as an empirical likelihood estimator for conditional moment restriction models.

Keywords: Conditional moment restrictions; Empirical likelihood; Fisher information; Least favorable submodel.
JEL classification: C14.

1 Introduction

Models specified via conditional moment restrictions are common in econometrics, and several estimators have been proposed in the literature. Contributors include Newey (1990), Carrasco and Florens (2000), Donald, Imbens, and Newey (2003), Zhang and Gijbels (2003), Dominguez and Lobato (2004), Kitamura, Tripathi, and Ahn (2004), and Lavergne and Patilea (2013).

The efficiency bound for conditional moment restriction models was originally derived by Chamberlain (1987), who utilized the fact that any distribution can be approximated arbitrarily well by a multinomial distribution. Severini and Tripathi (2001), who provided a general framework to derive the efficiency bound for semiparametric models using a Hilbert space theory, obtained the efficiency bound for conditional moment restriction models as a special case.

This study gives a simple and constructive method to derive the efficiency bound for conditional moment restriction models, based on the viewpoint of Stein (1956), who stated that the Fisher information in a semiparametric model is not larger than that in any parametric

*This research was supported by JSPS KAKENHI Grand Number 25780150.
†Graduate School of Economics, 2-1 Rokkodai-cho, Nada-ku, Kobe, Hyogo 657-8501, Japan. Email: sueishi@econ.kobe-u.ac.jp

submodel that satisfies the semiparametric assumption and contains the truth. A submodel is called least favorable if its Fisher information takes the infimum of the Fisher information over all smooth submodels, and the Fisher information in the semiparametric model is defined by that in the least favorable model. We calculate the Fisher information in conditional moment restriction models by deriving a least favorable model in an explicit form. Severini and Wong (1992) (hereafter: SW, 1992) showed that a least favorable submodel is obtained by finding the minimizer of the Kullback–Leibler discrepancy from the model to the true probability measure. Combining this result with a duality theorem in the convex analysis, we obtain the efficiency bound.

Furthermore, the proposed method suggests an asymptotically efficient estimator, which is essentially the same as the estimator of SW (1992) for conditionally parametric models. Moreover, our estimator can be viewed as an empirical likelihood (EL) estimator for conditional moment restriction models. We will observe a connection between our estimator and the smoothed EL (SEL) estimator of Kitamura, Tripathi, and Ahn (2004).

2 Fisher Information

Let $(X_1, Z_1), \ldots, (X_n, Z_n)$ be independent copies of (X, Z) that takes values in $\mathcal{X} \times \mathcal{Z}$. The joint law of (X, Z) is denoted by $\mu_{X,Z}$. Furthermore, we write $\mu_{X,Z} = \mu_{Z|X}\mu_X$, where $\mu_{Z|X}$ is the conditional law of Z given X and μ_X is the marginal law of X. The parameter of interest, $\theta_0 \in \Theta$, is characterized as a unique value that satisfies the conditional moment restriction

$$E[m(Z, \theta_0)|X] = \int m(z, \theta_0) d\mu_{Z|X} = 0 \quad a.s.\ \mu_X,$$

where $m : \mathcal{Z} \times \Theta \to \mathbb{R}^l$ is a known function. We assume that μ_X does not provide any information on θ_0. Moreover, without loss of generality, we assume that θ_0 is a scalar parameter. The model is semiparametric in the sense that $\mu_{Z|X}$ is the infinite-dimensional nuisance parameter. Our aim is to find the efficiency bound for estimating θ_0.

The model can be represented as a set of conditional probability measures $\{P_{\theta,\eta}\}$ that is indexed by the scalar parameter θ and the infinite-dimensional nuisance parameter η. We employ a setting that is used in Kitamura (2003) and Otsu and Whang (2011). Let $\mathcal{M}_{Z|x}$ be the set of conditional probability measures on $\mathcal{Z}$ given as $X = x$. Because a model for the marginal law of X is not specified, our model is written as

$$\mathcal{P}_{Z|X} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{H}_{\theta,X}\ a.s.\ \mu_X\},$$

where

$$\mathcal{H}_{\theta,x} = \left\{ \nu_{Z|x} \in \mathcal{M}_{Z|x} : \int m(z, \theta) d\nu_{Z|x} = 0 \right\}.$$

Note that $\mathcal{H}_{\theta,x}$ is the set of conditional probability measures that satisfies the conditional moment restriction for a given $\theta \in \Theta$ and $x \in \mathcal{X}$.

The Fisher information for θ_0 in $\mathcal{P}_{Z|X}$ is the infimum of Fisher information over all parametric submodels of $\mathcal{P}_{Z|X}$. Let η_0 be the true nuisance parameter that satisfies $\mu_{Z|X} = P_{\theta_0,\eta_0}$.

Moreover, let η_θ be a smooth curve of θ that satisfies $\eta_{\theta_0} = \eta_0$. Then we obtain a parametric submodel $\{P_{\theta,\eta_\theta} : \theta \in U\}$, where U is a neighborhood of θ_0. Fisher information in the submodel is

$$E\left[\left(\left.\frac{d\log dP_{\theta,\eta_\theta}}{d\theta}\right|_{\theta=\theta_0}\right)^2\right],$$

where the symbol E denotes the expectation with respect to $\mu_{X,Z}$. A curve η^*_θ that gives the infimum of the Fisher informations is called a least favorable curve. The model $\{P_{\theta,\eta^*_\theta} : \theta \in U\}$ is a least favorable submodel.

SW (1992) showed that a least favorable curve is obtained by minimizing the Kullback–Leibler discrepancy from $P_{\theta,\eta}$ to P_{θ_0,η_0} with respect to η. Thus, the least favorable curve is obtained by solving

$$\min_{\eta\in\mathcal{H}_{\theta,x}} -\int \log\left(\frac{dP_{\theta,\eta}}{d\mu_{Z|x}}\right)d\mu_{Z|x} = \min_{\eta\in\mathcal{H}_{\theta,x}} -\int \log\left(\frac{d\eta}{d\mu_{Z|x}}\right)d\mu_{Z|x}$$

for each θ and x. Moreover, a duality theorem implies that the infinite-dimensional minimization problem is equivalent to the finite-dimensional dual problem:

$$\max_{\lambda\in\mathbb{R}^l}\int \log(1+\lambda' m(z,\theta))d\mu_{Z|x}.$$

For instance, see Kitamura (2003).

The duality theorem implies that the least favorable submodel $\{P_{\theta,\eta^*_\theta}\}$ is equivalently written as $\{P_{\theta,\lambda_\theta}\}$ whose $\mu_{Z|x}$-density is given by

$$\frac{dP_{\theta,\lambda_\theta}}{d\mu_{Z|x}} = \frac{dP_{\theta,\eta^*_\theta}}{d\mu_{Z|x}} = (1+\lambda_\theta(x)' m(z,\theta))^{-1},$$

where

$$\lambda_\theta(x) = \arg\max_{\lambda\in\mathbb{R}^l}\int \log(1+\lambda' m(z,\theta))d\mu_{Z|x}.$$

The new nuisance parameter, λ, is finite-dimensional for fixed x.

The Fisher information in $\{P_{\theta,\lambda_\theta}\}$ is

$$E\left[\left(\left.\frac{d\ell(Z;\theta,\lambda_\theta(X))}{d\theta}\right|_{\theta=\theta_0}\right)\right] = E\left[\left(\frac{\partial\ell(Z;\theta_0,\lambda_{\theta_0}(X))}{\partial\theta} + \left[\frac{\partial\lambda_{\theta_0}(X)}{\partial\theta}\right]'\frac{\partial\ell(Z;\theta_0,\lambda_{\theta_0}(X))}{\partial\lambda}\right)^2\right],$$

where $\ell(z;\theta,\lambda) = -\log(1+\lambda' m(z,\theta))$. By the implicit function theorem, we have

$$\frac{\partial\lambda_\theta(x)}{\partial\theta} = -E\left[\frac{\partial^2\ell(Z;\theta,\lambda_\theta(X))}{\partial\lambda\partial\lambda'}|X=x\right]^{-1}E\left[\frac{\partial^2\ell(Z;\theta,\lambda_\theta(X))}{\partial\lambda\partial\theta}|X=x\right].$$

Because $\lambda_{\theta_0}(x) = 0$ for all $x\in\mathcal{X}$, some calculation yields that

$$\frac{\partial\lambda_{\theta_0}(x)}{\partial\theta} = -E\left[m(Z,\theta_0)m(Z,\theta_0)'|X=x\right]^{-1}E\left[\frac{\partial m(Z,\theta)}{\partial\theta}|X=x\right].$$

Thus, Fisher information in $\mathcal{P}_{Z|X}$ is

$$\mathcal{I}_{\theta_0} = E\left[E\left[\frac{\partial m(Z,\theta_0)}{\partial\theta}|X\right]'E\left[m(Z,\theta_0)m(Z,\theta_0)'|X\right]^{-1}E\left[\frac{\partial m(Z,\theta_0)}{\partial\theta}|X\right]\right].$$

The asymptotic variance of any regular estimator for θ_0 is not less than $\mathcal{I}^{-1}_{\theta_0}$.

3 Efficient Estimator

SW (1992) proposed an asymptotically efficient estimator for conditionally parametric models, in which a nuisance parameter is a Euclidean parameter once the value of a conditioning variable is fixed. Our original model is not conditionally parametric because η is infinite-dimensional even for a fixed value of X. However, the model in the dual problem is conditionally parametric, and thus the estimator of SW (1992) is applicable.

An infeasible efficient estimator for θ_0 is the maximum likelihood estimator that maximizes $\sum_{i=1}^{n} \ell(Z_i; \theta, \lambda_\theta(X_i))$. Because $\lambda_\theta(x)$ is unknown, we replace it with a sample analog. Suppose that $\mathcal{X}$ is compact. Then we can estimate $\lambda_\theta(x)$ by

$$\hat{\lambda}_\theta(x) = \arg\max_{\lambda \in \mathbb{R}^l} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) \ell(Z_j; \theta, \lambda),$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Our estimator for θ_0 is

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \ell(Z_i; \theta, \hat{\lambda}_\theta(X_i)),$$

which is essentially the same as the estimator of SW (1992).

Our estimator can also be viewed as an EL estimator. Indeed, if the model is specified via the unconditional moment restriction $E[m(Z, \theta_0)] = 0$, then a least favorable submodel is obtained as

$$\frac{dP_{\theta,\lambda_\theta}}{d\mu_Z} = (1 + \lambda_\theta' m(z, \theta))^{-1},$$

where

$$\lambda_\theta = \arg\max_{\lambda \in \mathbb{R}^l} \int \log(1 + \lambda' m(z, \theta)) d\mu_Z.$$

Thus, the resulting estimator coincides with the EL estimator of Qin and Lawless (1994). See also DiCiccio and Romano (1990) and Bertail (2006) for a connection between EL estimators and least favorable submodels.

Under certain regularity conditions, we have $\hat{\theta} \xrightarrow{p} \theta_0$ and

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}_{\theta_0}^{-1}).$$

Thus, the estimator is asymptotically efficient. Note that replacing λ_θ with $\hat{\lambda}_\theta$ does not affect the asymptotic distribution of the estimator. This is because λ_θ is a least favorable curve (see Section 4 of SW, 1992).

Because regularity conditions and proofs are given in SW (1992), we do not reproduce them. To show the asymptotic results, we need to establish uniform consistency of $\hat{\lambda}_\theta(x)$ and its derivatives over $\theta \in \Theta$ and $x \in \mathcal{X}$. Thus, if $\mathcal{X}$ is not compact, we need to introduce a trimming function to avoid the random denominator problem.

Finally, we note a connection between our estimator and the SEL estimator of Kitamura, Tripathi, and Ahn (2004). In our notation, the SEL estimator solves

$$\max_{\theta \in \Theta} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \ell(Z_j; \theta, \hat{\lambda}_\theta(X_i)),$$

where

$$w_{ij} = \frac{K\left(\frac{X_i - X_j}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{X_i - X_j}{h}\right)}.$$

Thus, the difference between two estimators is in whether the log-likelihood is smoothed when the "outer loop" is solved. The local likelihood method of Tibshirani and Hastie (1987) suggests that smoothing in the outer loop is necessary when θ_0 is a function of x, which is the case considered by Zhang and Gijbels (2003).

References

Bertail, P. (2006): "Empirical Likelihood in Some Semiparametric Models," *Bernoulli*, 12, 299–331.

Carrasco, M., and J.-P. Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

Chamberlain, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

DiCiccio, T. J., and J. P. Romano (1990): "Nonparametric Confidence Limits by Resampling Methods and Least Favorable Families," *International Statistical Review*, 58, 59–76.

Dominguez, M. A., and I. N. Lobato (2004): "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, 72, 1601–1615.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, 117, 55–93.

Kitamura, Y. (2003): "A Likelihood-Based Approach to the Analysis of a Class of Nested and Non-Nested Models," Unpublished Manuscript.

Kitamura, Y., G. Tripathi, and H. Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667–1714.

Lavergne, P., and V. Patilea (2013): "Smooth Minimum Distance Estimation and Testing with Conditional Estimating Equations: Uniform in Bandwidth Theory," *Journal of Econometrics*, 177, 47–59.

Newey, W. K. (1990): "Efficient Instrumental Variable Estmation of Nonlinear Models," *Econometrica*, 58, 809–837.

Otsu, T., and Y.-J. Whang (2011): "Testing for Nonnested Conditional Moment Restrictions via Condtional Empirical Likelihood," *Econometric Theory*, 27, 114–153.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and Generalized Estimating Equations," *Annals of Statistics*, 22, 300–325.

Severini, T. A., and G. Tripathi (2001): "A Simplified Approach to Computing Efficiency Bounds in Semiparametric Models," *Journal of Econometrics*, 102, 23–66.

Severini, T. A., and W. H. Wong (1992): "Profile Likelihood and Conditionally Parametric Models," *Annals of Statistics*, 20, 1768–1802.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Praobability*, vol. 1, pp. 187–195.

Tibshirani, R., and T. Hastie (1987): "Local Likelihood Estimation," *Journal of the American Statistical Association*, 82, 559–567.

Zhang, J., and I. Gijbels (2003): "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares," *Scandinavian Journal of Statistics*, 30, 1–24.