Minimum Sliced Distance Estimation in Structural Models

Hyeonseok Park

University of Washington

Kantorovich Initiative Retreat
March 18, 2022

Overview

- Structural Econometric Models and Obstacles in Estimation
- Parameter Estimation using Sliced Distances
    - Models and Estimators
    - Non-technical Discussions with Illustrative Examples
- Concluding Remarks

This talk is based on a joint work with Yanqin Fan.

Structural Econometric Models

$$\begin{matrix} X_t \text{ (Observable)} \\ \xi_t \text{ (Unobservable)} \end{matrix} \xrightarrow[\substack{\text{Economic Model} \\ Y_t = g(X_t, \xi_t, \theta)}]{} Y_t \text{ (Observable)}$$

Objective: To study relationships between economic variables using data.

Some examples:

- Understand the dynamics of stock returns (e.g, asset pricing models)
- Analyze the market of certain products such as automobile [Berry et al., 1995]
- Explain the observed black-white difference (e.g, wage) in the labor market [Bowlus et al., 2001]

Structural Econometric Models

The estimation of structural econometric models:

- Conventional Approach: Maximum Likelihood Estimation (MLE)
- The likelihood-based methods are not applicable for some cases.

Possible Obstacles in Estimation of Structural Models

- The likelihood function of observable variables may not exist.
  - For example, we might want to explain the dynamics of stock returns with a few unobservable (economic) factors.
- The likelihood function may be intractable even when it exists.
  - For example, the economic model could be too complicated.

Possible Obstacles in Estimation of Structural Models

- The support of an random variable might depend on the model parameters.
  - In particular, the likelihood function may be discontinuous depending on model parameters.
  - MLE may or may not asymptotically follow a normal distribution depending on model parameters.
  - The asymptotic theory of MLE may be sensitive to model assumptions.
  - However, it may be hard to verify whether the model assumptions hold.

Advantage of using OT in Structural Econometric Models

- Minimum distance estimator using the (sliced) L_2 distance between empirical and model-induced probability measures.
- Asymptotic normality of the estimator to a broad class of structural models.

The advantage is that estimators using sliced distances can be robust to model structures and assumptions.

Models

- Let $\{Z_t : \ t = 1, 2, \ldots\}$ be a d-dimensional stationary and weakly dependent time series.
- Let $F(\cdot; \theta_0)$ be the distribution function of Z_t where θ_0 is the parameter of interest.
- Given data $Z_1, \ldots, Z_T$, we wish to conduct inference on $\theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$.
- We estimate the parameter θ_0 in the model using the minimum distance estimator based on sliced distances.

Estimator using the Sliced Distance

The minimum (weighted) sliced Wasserstein distance (MSWD) estimator:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \int_{\mathbb{S}^{d-1}} \int_0^1 \left(G_T^{-1}(s; u) - G^{-1}(s; u, \theta) \right)^2 w(s) \, \mathrm{d}s \mathrm{d}\varsigma(u),$$

where $w(s)$ is a non-negative function such that $\int_0^1 w(s)\mathrm{d}s = 1$, and $\varsigma(u)$ is uniform distribution on the unit-sphere $\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : \|u\|_2 = 1\}$.
For each $u \in \mathbb{S}^{d-1}$,

$$G_T(s; u) = \frac{1}{T} \sum_{t=1}^{T} I(u^\top Z_t \leq s), \text{ and}$$

$$G(s; u, \theta) = \int I(u^\top z \leq s) \mathrm{d}F(z, \theta).$$

Illustrative Examples

- We use two examples to illustrate the different behaviors of MLE and the MSWD estimator.
  - A stochastic singular model.
  - The two-sided uniform model.
- In both examples, we compute 3000 estimates from the samples of size $T = 1000$.

Example 1.1.

- Let $Z = g_{\theta_0}(\xi) = (\theta_0, \xi)$, where $\theta_0 = 0$ and $\xi \sim U[0,1]$.
- The support of $g_{\theta_0}(\xi)$ is $\{\theta_0\} \times [0,1]$.
- (θ_0, ξ) and (θ, ξ) have disjoint supports unless $\theta = \theta_0$.

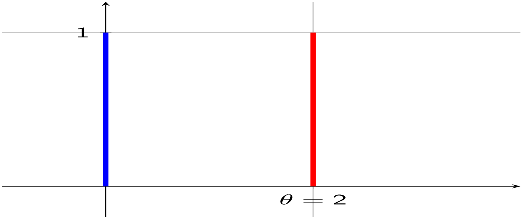

Figure 1: The supports of (θ, ξ) for $\theta = 0$ (blue) and $\theta = 2$ (red)

Numerical Illustration

A stochastic singular model in Arjovsky et al. [2017]

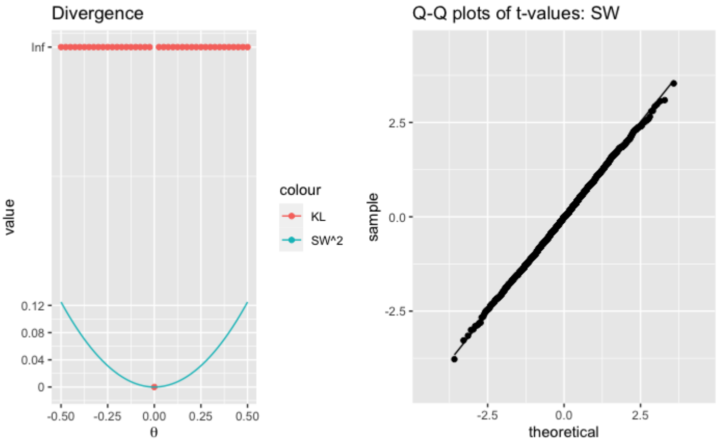

Figure 2: A comparison of several divergences when $T = 1000$

Example 1.2.

Suppose the support of Z is $[0, 1]$ and its density function is

$$f(z;\theta_0) = \begin{cases} 0.25\theta_0^{-1} & \text{if } 0 \leq z \leq \theta_0, \\ 0.75(1-\theta_0)^{-1} & \text{if } \theta_0 < z \leq 1. \end{cases}$$

- The density function has a jump at θ_0 with jump size $\frac{4\theta_0 - 1}{4\theta_0(1-\theta_0)}$.
- The jump size is zero when $\theta_0 = 1/4$ and non-zero otherwise.

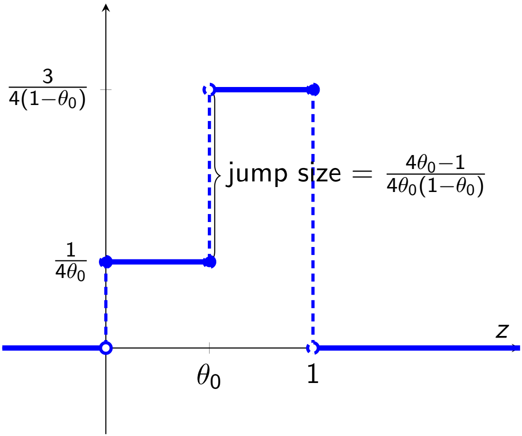

Figure 3: The density function of two-sided uniform distribution

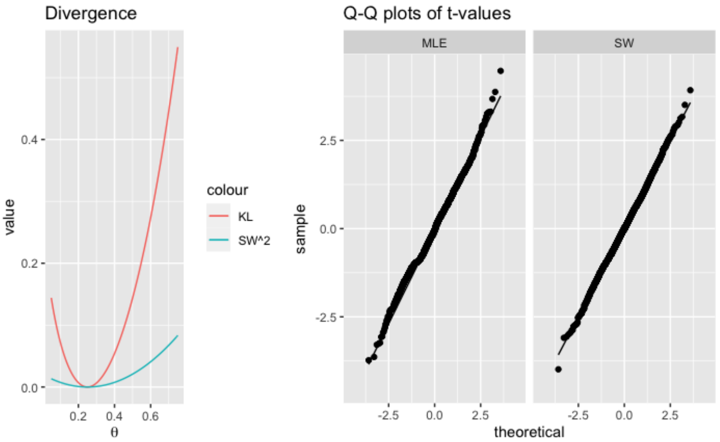

Figure 4: A comparison of several divergences when $\theta_0 = 1/4$ with $T = 1000$

Numerical Illustration

Two-sided Uniform Model

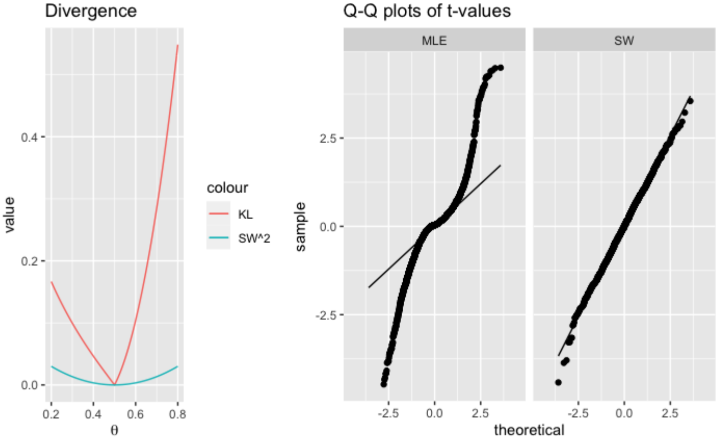

Figure 5: A comparison of several divergences when $\theta_0 = \mathbf{1/2}$ with $T = 1000$

Nontechnical Discussions

The minimum (weighted) sliced Wasserstein distance (MSWD) estimator:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \int_{\mathbb{S}^{d-1}} \int_0^1 \left(G_T^{-1}(s; u) - G^{-1}(s; u, \theta) \right)^2 w(s) \, \mathrm{d}s \mathrm{d}\varsigma(u),$$

- When $G^{-1}(s, u, \theta)$ is sufficiently smooth in θ, the estimator can achieve the asymptotic normality under some regularity conditions.

tech-assumptions

Nontechnical Discussions

The minimum (weighted) sliced Wasserstein distance (MSWD) estimator:

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} \int_{\mathbb{S}^{d-1}} \int_0^1 \left(G_T^{-1}(s; u) - G^{-1}(s; u, \theta) \right)^2 w(s) \, \mathrm{d}s \mathrm{d}\varsigma(u),$$

- Why this approach can be robust in the structural econometric models?
  - Even when the model is singular, the distribution function of one-dimensional projection $u^\top Z_t$ can be well-defined.
  - The distribution function might be smoother than the density function.
  - When $G(s, u)$ is not feasible, we might construct empirical distribution by generating thy synthetic data from the model.

Concluding Remarks

- The estimation using the sliced (Wasserstein distance) can provide a simple and robust method for estimation of structural econometric models.
- Important issues remain to be addressed.
  - More sophisticated computational algorithms for the complex models such as the demand models.
  - Extension to possibly misspecified models.

References I

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 10 2017.

David K. Backus, Silverio Foresi, and Chris Telmer. Discrete-time models of bond pricing. *NBER Working Paper*, (6736), 1998.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile Prices in Market Equilibrium. *Econometrica*, 63(4):841, 1995. ISSN 0012-9682. doi: 10.2307/2171802.

Sergey Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Memoirs of the American Mathematical Society*, 261(1259):0, 2019. doi: 10.1090/memo/1259.

References II

Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and Radon Wasserstein Barycenters of Measures. *Journal of Mathematical Imaging and Vision*, 51(1), 1 2015. ISSN 0924-9907. doi: 10.1007/s10851-014-0506-3.

Audra J. Bowlus, Nicholas M. Kiefer, and George R. Neumann. Equilibrium search models and the transition from school to work. *International Economic Review*, 42(2), 2001. ISSN 00206598. doi: 10.1111/1468-2354.00112.

Richard C Bradley. Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. *Probability Surveys*, 2(1):107–144, 2005. doi: 10.1214/154957805100000104.

Donald J. Brown and Marten H. Wegkamp. Weighted minimum mean-square distance from independence estimation, 2002. ISSN 00129682.

References III

Tetsuya Kaji, Elena Manresa, and Guillaume Pouliot. An Adversarial Approach to Structural Estimation. *SSRN Electronic Journal*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3706365.

Sergio Pastorello, Valentin Patilea, and Eric Renault. Iterative and Recursive Estimation in Structural Nonadaptive Models. *Journal of Business & Economic Statistics*, 21(4):449–509, 10 2003. ISSN 0735-0015. doi: 10.1198/073500103288619124.

Alexander Torgovitsky. Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics*, 199(1), 2017. ISSN 18726895. doi: 10.1016/j.jeconom.2017.01.009.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer New York, New York, NY, 1996. ISBN 978-1-4757-2547-6. doi: 10.1007/978-1-4757-2545-2.

General Estimator

Given the sample information $\{Z_t\}_{t=1}^{T}$,

- let $Q_T(\cdot; u)$ denote an empirical measure such as the empirical quantile or empirical distribution function of $\left\{u^\top Z_t\right\}_{t=1}^{T}$;
- $\widehat{Q}_T(\cdot; u, \theta)$ denote a possibly random function depending on the model, $\theta \in \theta \subset \mathbb{R}^{d_\theta}$.
  - In unconditional models, it is the parametric quantile or distribution function of $u^\top Z_t$.
  - In generative models, it could be an empirical measure using generated data.

non-tech-discussion

General Estimator

A general minimum sliced distance (MSD hereafter) estimator denoted by $\hat{\theta}_T$ is defined as

$$\hat{\theta}_T = \operatorname*{argmin}_{\theta \in \Theta} \int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} (Q_T(s; u) - \widehat{Q}_T(s; u, \theta))^2 w(s) \mathrm{d}s d\varsigma(u), \qquad (1)$$

where $\mathcal{S}$ is the domain of $Q_T(s; u)$ with respect to s.

For the MSWD estimator of unconditional models,

- $Q_T(s; u)$ is the empirical quantile function of $\{u^\top Z_t\}_{t=1}^T$.
- $\widehat{Q}_T(s; u, \theta)$ is the parametric quantile function of $u^\top Z_t$.

General Asymptotic Theory (Consistency of $\hat{\theta}$)

Assumption 8.1.

(i) $\int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} (Q_T(s;u) - Q(s;u))^2 w(s) \mathrm{d}s d\varsigma(u) \xrightarrow{p} 0.$

(ii) $\sup_{\theta \in \theta} \int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} \left(\widehat{Q}_T(s;u,\theta) - Q(s;u,\theta) \right)^2 w(s) \mathrm{d}s d\varsigma(u) \xrightarrow{p} 0.$

- $Q(\cdot,;u)$ represent a probability measure such as parametric quantile or distribution function of $u^\top Z_t$, where $Z_t \sim F(\cdot, \theta_0)$, and
- $Q(\cdot,;u,\theta)$ represent a probability measure such as parametric quantile or distribution function of $u^\top Z_t$, where $Z_t \sim F(\cdot, \theta)$.

General Asymptotic Theory (Consistency of $\hat{\theta}_T$)

Assumption 8.2 (Identification).

The parameter θ_0 is in the interior of θ such that for all $\epsilon > 0$, and

$$\inf_{\theta \notin B(\theta_0, \epsilon)} \int_{\mathbb{S}^{d-1}} \int_S (Q(s; u) - Q(s; u, \theta))^2 w(s) \mathrm{d}s \mathrm{d}\varsigma(u) > 0,$$

where $B(\theta_0, \epsilon) := \{\theta \in \theta : \|\theta - \theta_0\| \leq \epsilon\}$.

Theorem 8.1 (Consistency of $\hat{\theta}_T$).

Suppose Assumptions 8.1 and 8.2 hold. Then $\hat{\theta}_T \xrightarrow{p} \theta_0$ as $T \to \infty$.

General Asymptotic Theory (Asymptotic Normality of $\hat{\theta}$)

Let

$$\widehat{R}_T(s; u, \theta, \theta_0) := \widehat{Q}_T(s; u, \theta) - \widehat{Q}_T(s; u, \theta_0) - (\theta - \theta_0)^\top \widehat{D}_T(s; u, \theta_0),$$

where $\widehat{D}_T(\cdot\,; \cdot,\, \theta_0)$ is an $L^2(\mathcal{S} \times \mathbb{S}^{d-1}, w(s)\mathrm{d}s\mathrm{d}\varsigma(u))$-measurable function.

Assumption 8.3 (Norm-differentiability).

$\widehat{Q}_T(\cdot; u, \theta)$ *is first-order norm-differentiable at* $\theta = \theta_0$. *That is,*

$$\sup_{\theta \in \theta;\ \|\theta - \theta_0\| \leq \tau_T} \left| \frac{T \int_{\mathbb{S}^{d-1}} \int_S \left(\widehat{R}_T(s; u, \theta, \theta_0) \right)^2 w(s) ds d\varsigma(u)}{(1 + \|\sqrt{T}(\theta - \theta_0)\|)^2} \right| = o_p(1)$$

for any $\tau_T \to 0$.

Assumption 8.4.

The following conditions hold:

(i) $T \int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} (Q_T(s;u) - Q(s;u))^2 w(s) \, \mathrm{d}s d\varsigma(u) = O_p(1)$;

(ii) $T \int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} (\widehat{Q}_T(s;u,\theta_0) - Q(s;u,\theta_0))^2 w(s) \mathrm{d}s d\varsigma(u) = O_p(1)$;

(iii) *There exists an* $L^2(\mathbb{R} \times \mathbb{S}^{d-1}, w(s)\mathrm{d}s\mathrm{d}\varsigma)$*-measurable function* $D(\cdot;\cdot,\theta_0)$ *such that*

$$\int_{\mathbb{S}^{d-1}} \int_{\mathcal{S}} \left\| \widehat{D}_T(s;u,\theta_0) - D(s;u,\theta_0) \right\|^2 w(s) \mathrm{d}s d\varsigma(u) = o_p(1).$$

- Assumption 8.4 (i) strengthens Assumption 8.1 (i).

General Asymptotic Theory (Asymptotic Normality of $\hat{\theta}$)

Assumption 8.5.

$$\sqrt{T}\begin{pmatrix} \int_{\mathbb{S}^{d-1}}\int_{\mathcal{S}}(Q_T(s;u)-Q(s;u))D(s;u.\theta_0)w(s)dsd\varsigma(u) \\ \int_{\mathbb{S}^{d-1}}\int_{\mathcal{S}}(\widehat{Q}_T(s;u,\theta_0)-Q(s;u,\theta_0))D(s;u,\theta_0)w(s)dsd\varsigma(u) \end{pmatrix}$$

$\xrightarrow{d} N(0,V_0)$ *for some positive semidefinite matrix* V_0.

Assumption 8.6.

The matrix

$$B_0 := \int_{\mathbb{S}^{d-1}}\int_{\mathcal{S}} D(s;u,\theta_0)D^\top(s;u,\theta_0)w(s)\mathrm{d}sd\varsigma(u) \tag{2}$$

is positive definite.

General Asymptotic Theory (Asymptotic Normality of $\hat{\theta}$)

Theorem 8.2 (Asymptotic normality of $\hat{\theta}_T$).

Suppose Assumptions 8.1 to 8.6 hold. Then,

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, B_0^{-1}\Omega_0 B_0^{-1}),$$

where $\Omega_0 = (e_1^\top, -e_1^\top) V_0 \begin{pmatrix} e_1 \\ -e_1 \end{pmatrix}$ *in which* $e_1 = (1, \dots, 1)^\top$ *is a* d_θ*-dimensional vector of ones.*