Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 09-012

"Two-Step Extremum Estimation with Estimated Single-Indices"

by

Kyungchul Song

http://ssrn.com/abstract=1360975

Two-Step Extremum Estimation with Estimated Single-Indices

Kyungchul Song[1]
Department of Economics, University of Pennsylvania

February 16, 2009

Abstract

This paper studies two-step extremum estimation that involves the first step estimation of nonparametric functions of single-indices. First, this paper finds that under certain regularity conditions for conditional measures, linear functionals of conditional expectations are insensitive to the first order perturbation of the parameters in the conditioning variable. Applying this result to symmetrized nearest neighborhood estimation of the nonparametric functions, this paper shows that the influence of the estimated single-indices on the estimator of main interest is asymptotically negligible even when the estimated single-indices follow cube root asymptotics. As a practical use of this finding, this paper proposes a bootstrap method for conditional moment restrictions that are asymptotically valid in the presence of cube root-converging single-index estimators. Some results from Monte Carlo simulations are presented and discussed.

Keywords: two-step extremum estimation; single-index restrictions; cube root asymptotics; bootstrap;

JEL Classifications: C12, C14, C51.

1 Introduction

Many empirical studies use a number of covariates to deal with the problem of endogeneity. Using too many covariates in nonparametric estimation, however, tends to worsen the quality of the empirical results significantly. A promising approach in this situation is to introduce a

[1] I thank Xiaohong Chen, Stefan Hoderlein, Simon Lee, Frank Schorfheide and seminar participants at the Greater New York Econometrics Colloquium at Princeton University for valuable comments. All errors are mine. Address correspondence to Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Bldg, 3718 Locust Walk, Philadelphia, PA 19104-6297.

single-index restriction so that one can retain flexible specification while avoiding the curse of dimensionality. The single-index restriction has long attracted attention in the literature. For example, Klein and Spady (1993) and Ichimura (1993) proposed M-estimation approaches to estimate the single-index, and Stoker (1986) and Powell, Stock and Stoker (1989) proposed estimation based on average derivatives. See also Härdle and Tsybakov (1993), Härdle, Hall and Ichimura (1993), Horowitz and Härdle (1996), and Hristache, Juditsky and Spokoiny (2001).

Most literatures have dealt with a single-index model as an isolated object, whereas researchers often use it as part of a larger model. This paper considers the following estimation framework. Let the parameter of interest $\beta_0 \in \mathbf{R}^d$ be identified as the unique maximizer of a population objective function :

$$\beta_0 = \text{argmax}_\beta Q(\beta, \mu_0(\cdot; \lambda_0)), \tag{1}$$

where $\mu_0(\cdot; \lambda_0) = (\mu_{0,1}(\cdot; \lambda_{0,1}), \cdots, \mu_{0,J}(\cdot; \lambda_{0,J}))^\top$ and

$$\mu_{0,j}(\cdot; \lambda_{0,j}) = \mathbf{E}[Y^{(j)} | \lambda_{0,j}(X) = \lambda_{0,j}(\cdot)]$$

with $Y^{(j)}$ being the j-th component of random vector $Y \in \mathbf{R}^J$ and X being a random vector in $\mathbf{R}^{d_X}$. The real function $\lambda_{0,j} : \mathbf{R}^{d_X} \to \mathbf{R}$ is a single-index of X. The distributions of $\lambda_{0,j}(X)$'s are assumed to be absolutely continuous.

We assume that μ_0 and λ_0 are identified and estimated prior to estimating β_0. The identification is ensured either through a single-index restriction imposed on an identified nonparametric function or through some auxiliary data set in the sense of Chen, Hong, and Tarozzi (2008). Then the estimator of β_0 can be constructed as

$$\hat{\beta} = \text{argmax}_\beta Q_n(\beta, \hat{\mu}(\cdot; \hat{\lambda})), \tag{2}$$

where $Q_n(\beta, \hat{\mu}(\cdot; \hat{\lambda}))$ is the sample objective function and $\hat{\mu}(\cdot; \hat{\lambda})$ is the nonparametric estimator of $\mu_0(\cdot; \lambda_0)$ using $\hat{\lambda}$, an estimator of λ_0. The function $\lambda_{0,j}$ is either a nonparametric function or a parametric function. In the latter case, the estimator $\hat{\lambda}_j$ is allowed to be either $\sqrt{n}$-consistent or $n^{1/3}$-consistent.

The main finding of this paper is that there is no estimation effect of $\hat{\lambda}$ upon the asymptotic variance matrix of $\hat{\beta}$ under certain regularity conditions. (See Theorem 1 below.) Newey (1994) explained how the first step estimators affect the asymptotic variance of the second step estimators. The influence of the first step estimators is represented through a pathwise derivative of the parameter of interest in the nuisance parameters. However, the nature of the problem here is different in the sense that the nonparametric function $\mu_0(\cdot;\lambda_0)$ depends on λ_0 through the σ-field generated by $\lambda_0(X)$. Therefore, it is not immediately obvious to find the pathwise derivative of the parameter in λ_0. Note also that the usual analysis through an asymptotic linear representation of $\hat{\lambda}$ does not help either when $\hat{\lambda}$ follows cube root asymptotics because such a linear representation does not exist in this case.

First, the paper introduces regularity conditions for conditional measures and show that under these conditions, linear functionals of $\mu_0(\cdot;\lambda)$ have a zero Fréchet derivative in λ (Lemma 2). Using this result, the paper establishes a uniform Bahadur representation of sample linear functionals of the symmetrized nearest neighborhood (SNN) estimator (Lemma A1 in the Appendix). Through the uniform representation, it is shown that there is no estimation effect of $\hat{\lambda}$ upon the asymptotic variance of $\hat{\beta}$.

The asymptotic negligibility of the estimated single-index has broad implications for inference of various semiparametric models. Among other things, the result of this paper illuminates the asymptotic theory of estimators from certain models that have not appeared in the literature. Examples are a sample selection model with conditional median restrictions and models with single-index instrumental variables that are estimable at the rate of $n^{1/3}$. Second, there can be valid bootstrap methods for the inference of β_0 even when $\hat{\lambda}$ follows cube root asymptotics. This is interesting because bootstrap is known to fail for such $n^{1/3}$-converging estimators (Abrevaya and Huang (2005).) This paper proposes a bootstrap method in the special case of conditional moment restrictions.

A similar finding for $\sqrt{n}$-consistent single-index estimators has already appeared in Fan and Li (1996) in the context of testing semiparametric models. See also Stute and Zhu (2005) for a related result in testing single-index restrictions. These literatures deal with a special case where the single-index component is a parametric function with a $\sqrt{n}$-consistent estimator. This paper places in the broad perspective of extremum estimation the phenomenon of asymptotic negligibility of the estimated single-index and allows for the single-index estimator to be a $n^{1/3}$-consistent estimator or a nonparametric estimator. Let us conclude the introduction by discussing some examples.

Example 1 (Sample Selection Model with a Median Restriction) : Consider the following model:

$$\begin{aligned} Y &= \beta_0^\top W_1 + v \text{ and} \\ D &= 1\{\lambda_0(X) \geq \varepsilon\}, \end{aligned}$$

where $\lambda_0(X) = X^\top \theta_0$. The variable Y denotes the latent outcome and W_1 a vector of covariates that affect the outcome. The binary D represents the selection of the vector

(Y, W_1) into the observed data set, so that (Y, W_1) is observed only when $D = 1$. The incidence of selection is governed by a single index $\lambda_0(X)$ of covariates X. The variables v and ε represent unobserved heterogeneity in the individual observation.

The variable ε is permitted to be correlated with X but $Med(\varepsilon|X) = 0$. And W_1 is independent of (v, ε) conditional on the index $\lambda_0(X)$ in the selection mechanism. Therefore, the individual components of X can be correlated with v. The assumptions of the model are certainly weaker than the common requirement that (W_1, X) be independent of (v, ε). (e.g. Heckman (1990), Newey, Powell, and Walker (1990).) More importantly, this model does not assume that X is independent of unobserved component ε in the selection equation. Hence we cannot use the characterization of the selection bias through the propensity score $P\{D = 1|\lambda_0(X)\}$ as has often been done in the literature of semiparametric extension of the sample selection model. (e.g. Powell (1989), Ahn and Powell (1993), Chen and Khan (2003), and Das, Newey and Vella (2003)).

From the method of Robinson (1988), the identification of β_0 still follows if the matrix

$$\mathbf{E}\left[(X - \mathbf{E}[X|D = 1, \lambda_0(X)])(X - \mathbf{E}[X|D = 1, \lambda_0(X)])^\top |D = 1\right]$$

is positive definite. In this case, we can write for the observed data set $(D = 1)$

$$Y = \beta_0^\top W_1 + \tau(\lambda_0(X)) + u,$$

where u satisfies that $\mathbf{E}[u|D = 1, W_1, \lambda_0(X)] = 0$ and τ is an unknown nonparametric function. This model can be estimated by using the method of Robinson (1988). Let $\mu_Y(\cdot) = \mathbf{E}[Y|D = 1, \lambda_0(X) = \cdot]$, and $\mu_{W_1}(\cdot) = \mathbf{E}[W_1|D = 1, \lambda_0(X) = \cdot]$. Then, we consider a conditional moment restriction:

$$\mathbf{E}\left[\{Y - \mu_Y(\lambda_0(X))\} - \beta_0^\top\{W_1 - \mu_{W_1}(\lambda_0(X))\}|D = 1, W_1, \lambda_0(X)\right] = 0.$$

One may estimate θ_0 in λ_0 using maximum score estimation in the first step and use it in the second step estimation of β_0. Then the remaining question centers on the effect of the first step estimator of θ_0 which follows cube root asymptotics upon the estimator of β_0.

Note that the identification of θ_0 does not stem from a direct imposition of single-index restrictions on $\mathbf{E}[Y|D = 1, X = \cdot]$ and $\mathbf{E}[Z|D = 1, X = \cdot]$. The identification follows from the use of auxiliary data set $((D = 0), X)$ in the sense of Chen, Hong, and Tarozzi (2008). Such a model of "single-index selectivity bias" has a merit of avoiding a strong exclusion restriction and has early precedents. See Powell (1989), Newey, Powell, and Walk (1990), and Ahn and Powell (1993). ■

Example 2 (Models with a Single-Index Instrumental Variable) : Consider the following model:

$$\begin{aligned} Y &= Z^\top \beta_0 + \varepsilon, \text{ and} \\ D &= 1\{\lambda_0(X) \geq \eta\}, \end{aligned}$$

where $\lambda_0(X) = X^\top \theta_0$ and ε and η satisfy that $\mathbf{E}[\varepsilon|\lambda_0(X)] = 0$ and $Med(\eta|X) = 0$. Therefore, the index $\lambda_0(X)$ plays the role of the instrumental variable (IV). However, the IV exogeneity condition is weaker than the conventional one because the exogeneity is required only of the single-index $X^\top \theta_0$ not the whole vector X. In other words, some of the elements of the vector X are allowed to be correlated with ε. Furthermore, X is not required to be independent of η as long as it maintains the conditional median restriction. This conditional median restriction enables one to identify θ_0 and in consequence β_0. Hence the data set (D, X) plays the role of an auxiliary data set in Chen, Hong, and Tarozzi (2008).

While there are many ways to estimate β_0, we consider the following conditional moment restriction:

$$\mathbf{E}\left[Y - \mathbf{E}[Z|\lambda_0(X)]^\top \beta_0 | \lambda_0(X)\right] = 0.$$

We can first estimate λ_0 and $\mathbf{E}[Z|\lambda_0(X)]$ and then estimate β_0 by plugging in these estimates into a sample version of the conditional moment restriction. ■

Example 3 (Models with Single-Index Restrictions) : There are numerous semiparametric models that contain nonparametric estimation of a function $\mathbf{E}[Y|X]$ in the first step. (e.g. Ahn and Manski (1993), Buchinsky and Hahn (1998), Hirano, Imbens, and Ridder (2003).) The finding of this paper enables one to employ the same asymptotic analysis in the literature when one imposes a single index restriction:

$$\mathbf{E}[Y|X] = m(X^\top \gamma_0)$$

for some unknown function m and parameter γ_0. We can estimate γ_0 using the methods of inference for single-index models and plug the estimator $\hat{\gamma}_0$ in the nonparametric estimation of m. The coefficient estimator $\hat{\gamma}$ is typically $\sqrt{n}$-consistent. Then the asymptotic analysis can be done as if we know the true index parameter γ_0, because the estimation error in $\hat{\gamma}_0$ does not affect the asymptotic variance of the parameter of interest. ■

Some models where an unknown nonparametric function $\lambda_0(\cdot)$ constitutes the conditioning variable of a conditional expectation have received attention in the literature.

Example 4 (Matching Estimators of Treatment Effects on the Treated) : Let Y_1

and Y_0 be potential outcomes of a treated and an untreated individuals and D the treatment status. The parameter of interest is $\mu_1 = \mathbf{E}[Y_1 - Y_0|D = 1]$, i.e., the treatment effect on the treated. Let $\lambda_0(X) = P\{D = 1|X\}$, where X is a vector of covariates. Under the condition:

$$\mathbf{E}[Y_0|\lambda_0(X), D = 0] = \mathbf{E}[Y_0|\lambda_0(X), D = 1], \tag{3}$$

we can identify (Heckman, Ichimura, and Todd (1997))

$$\mu_1 = \mathbf{E}\left[Y_1 - \mathbf{E}[Y_0|D = 0, \lambda_0(X)]|D = 1\right].$$

Therefore, the parameter of interest μ_1 involves a nonparametric function λ_0 in the conditioning variable. Then, following Heckman, Ichimura and Todd (1998), we can estimate μ_1 by

$$\hat{\mu}_1 = \frac{1}{\sum_{i=1}^n 1\{D_i = 1\}} \sum_{i=1}^n 1\{D_i = 1\} \left\{Y_{1i} - \hat{\mathbf{E}}[Y_{0i}|\hat{\lambda}(X_i), D = 0]\right\}, \tag{4}$$

where $\hat{\mathbf{E}}[Y_{0i}|\lambda(X_i), D = 0]$ is a nonparametric estimator of $\mathbf{E}[Y_{0i}|\lambda_0(X_i), D = 0]$ and $\hat{\lambda}(X)$ that of $\lambda_0(X)$. Therefore, it is important for the asymptotic variance of $\hat{\mu}_1$ to analyze the effect of estimation $\hat{\lambda}$. ■

The remainder of the paper has three sections. The first section exposits the main result of this paper and provides heuristics. The second section focuses on the case with conditional moment restrictions and proposes a valid bootstrap procedure in the presence of $n^{1/3}$-converging nuisance parameter estimators. The third section presents and discusses simulation results and the last section concludes. The appendix contains technical proofs of the main results and a general uniform Bahadur representation of sample linear functionals of SNN estimators.

2 The Main Results

2.1 A Motivating Example

To illustrate the main motivation of this paper, we present some simulation results from the following semiparametric model:

$$Y_i = Z_i\beta_0 + \gamma_0 f(X_i^\top\theta_0) + \varepsilon_i,$$

where $f(v)$ is unknown, $\mathbf{E}[\varepsilon_i|X_i^\top\theta_0, Z_i] = 0$ and θ_0 is identified and estimated using some other data sources. We first generated the following fictitious first step "estimator" with varied noise levels:

$$\tilde{\theta}_k = \theta_{0,k} + a \times N(0,1),\ \ k = 1, 2,$$

with $a \in \{0.2, 0.4, 0.6, 1, 2, 3, 4\}$. We normalized the scale and defined $\hat{\theta} = \tilde{\theta}/||\tilde{\theta}||$ as the first step "estimator" of θ_0. Using Robinson's procedure, we can write the model as a semiparametric conditional moment restriction. Then, in the second step, we estimated β_0 from this restriction. (Details are found in Section 3.)

The data generating process used is as follows. We drew $\varepsilon_i,\ v_i,\ w_i,\ \varepsilon_{1,i}$ and $\varepsilon_{2,i}$ independently from $N(0,1)$ and defined

$$Z_i = v_i + w_i,\ \text{ and } X_{k,i} = v_i + \varepsilon_{k,i},\ \ k = 1, 2.$$

We set $\theta_0 = [-0.5,\ 1]^\top$, $\gamma_0 = 0$, and $\beta_0 = 2$. The sample size was $n = 300$ and the Monte Carlo simulation number was 1000.

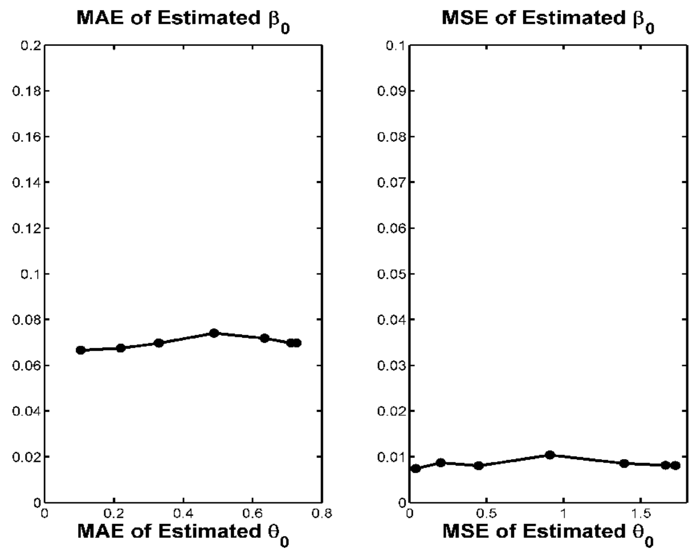
The results are shown in Figure 1 which plots the mean absolute error (MAE) and the mean squared error (MSE) of $\hat{\beta}$ against those of $\hat{\theta}$. The different points in the line represent results corresponding to the different choices of the noise level a. The results show that the quality of $\hat{\beta}$ is robust to that of $\hat{\theta}$, both in terms of MAE and MSE. The robustness of MSE of $\hat{\beta}$ against that of $\hat{\theta}$ is remarkable. This paper analyzes this phenomenon and reveals that it has a generic nature in a much broader context of extremum estimation. In particular, this robustness enables us to bootstrap $\hat{\beta}$ validly even when $\hat{\theta}$ follows cube root asymptotics in models of conditional moment restrictions.

2.2 Continuity of Linear Functionals of Conditional Expectations

Conditional expectations that involve unknown parameters in the conditioning variable frequently arise in semiparametric models. Continuity of conditional expectations with respect to such parameters plays a central role in this paper. In this section, we provide a generic, primitive condition that yields such continuity. Let $X \in \mathbf{R}^{d_X}$ be a random vector with support $\mathcal{S}_X$ and let Λ be a class of $\mathbf{R}$-valued functions on $\mathbf{R}^{d_X}$ with a generic element denoted by λ.

Fix $\lambda_0 \in \Lambda$ and let $f_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2)$ denote the conditional density function of a random vector $Y \in \mathbf{R}^{d_Y}$ given $(\lambda_0(X), \lambda(X)) = (\bar{\lambda}_1, \bar{\lambda}_2)$ with respect to a σ-finite measure, say, $w_\lambda(\cdot|\bar{\lambda}_1, \bar{\lambda}_2)$. Note that we do not assume that Y is absolutely continuous as we do not require that $w_\lambda(\cdot|\bar{\lambda}_1, \bar{\lambda}_2)$ is a Lebesgue measure. Let $\mathcal{S}_Y$ be the support of Y and let $\mathcal{S}_\lambda$ be that of $(\lambda_0(X), \lambda(X))$. We define $||\cdot||$ to be the Euclidean norm in $\mathbf{R}^J$ and $||\cdot||_\infty$ to be the sup norm: $||f||_\infty = \sup_{x\in\mathcal{S}_X}|f(x)|$.

Figure 1: The Robustness of the Second Step Estimator



Definition 1 : (i) $\mathcal{P}_Y \equiv \{f_\lambda(y|\cdot,\cdot) : (\lambda, y) \in \Lambda \times \mathcal{S}_Y\}$ is *regular for* $\tilde{\varphi} : \mathbf{R}^{d_Y} \to \mathbf{R}^J$, if for each $\lambda \in \Lambda$ and $(\bar{\lambda}_1, \bar{\lambda}_2) \in \mathcal{S}_\lambda$,

$$\sup_{(\tilde{\lambda}_1,\tilde{\lambda}_2)\in\mathcal{S}_\lambda:||\bar{\lambda}_1-\tilde{\lambda}_1||+||\bar{\lambda}_2-\tilde{\lambda}_2||\leq\delta} \left| f_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2) - f_\lambda(y|\tilde{\lambda}_1, \tilde{\lambda}_2) \right| < C_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2)\delta, \ \delta \in [0, \infty)$$

where $C_\lambda(\cdot|\bar{\lambda}_1, \bar{\lambda}_2) : \mathcal{S}_Y \to \mathbf{R}$ is such that for some $C > 0$,

$$\sup_{(y,\bar{\lambda}_1,\bar{\lambda}_2)\in\mathcal{S}_Y\times\mathcal{S}_\lambda} \int ||\tilde{\varphi}(y)||C_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2)w_\lambda(dy|\bar{\lambda}_1, \bar{\lambda}_2) < C.$$

(ii) When $\mathcal{P}_Y$ is regular for an identity map, we say simply that it is *regular*.

The regularity condition is a type of an equicontinuity condition for functions $f_\lambda(y|\cdot,\cdot)$, $(y, \lambda) \in \mathcal{S}_Y \times \Lambda$. Note that the condition does not require that the conditional density function be continuous in $\lambda \in \Lambda$, which is cumbersome to check in many situations. When $f_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2)$ is continuously differentiable in $(\bar{\lambda}_1, \bar{\lambda}_2)$ with a derivative that is bounded uniformly over $\lambda \in \Lambda$ and $\tilde{\varphi}(Y)$ has a bounded support, $\mathcal{P}_Y$ is regular for $\tilde{\varphi}$. Alternatively suppose that there exists $C > 0$ such that for each $\lambda \in \Lambda$ and $(\bar{\lambda}_1, \bar{\lambda}_2) \in \mathcal{S}_\lambda$,

$$\sup_{(\tilde{\lambda}_1,\tilde{\lambda}_2)\in\mathcal{S}_\lambda:||\bar{\lambda}_1-\tilde{\lambda}_1||+||\bar{\lambda}_2-\tilde{\lambda}_2||\leq\delta} \left| \frac{f_\lambda(y|\tilde{\lambda}_1, \tilde{\lambda}_2)}{f_\lambda(y|\bar{\lambda}_1, \bar{\lambda}_2)} - 1 \right| \leq C\delta,$$

and $\mathbf{E}[||\tilde{\varphi}(Y)|||X] < C$. Then $\mathcal{P}_Y$ is regular for $\tilde{\varphi}$. The regularity condition for $\mathcal{P}_Y$ yields the following important consequence. Define

$$\mu_\varphi(x;\lambda) = \mathbf{E}[\varphi(Y)|\lambda(X) = \lambda(x)],$$

where $\varphi \in \Phi$ with Φ being a class of $\mathbf{R}^J$-valued functions on $\mathbf{R}^{d_Y}$.

Lemma 1 : *Suppose that $\mathcal{P}_Y$ is regular for $\tilde{\varphi}$ an envelope of Φ. Then, for each $\lambda \in \Lambda$ and $x \in \mathcal{S}_X$,*

$$\begin{aligned} ||\mu_\varphi(x;\lambda_0,\lambda) - \mu_\varphi(x;\lambda)|| &\leq C|\lambda(x) - \lambda_0(x)|, \text{ and} \\ ||\mu_\varphi(x;\lambda_0,\lambda) - \mu_\varphi(x;\lambda_0)|| &\leq C|\lambda(x) - \lambda_0(x)|, \end{aligned}$$

where $\mu_\varphi(x;\lambda_0,\lambda) = \mathbf{E}[\varphi(Y)|(\lambda_0(X),\lambda(X)) = (\lambda_0(x),\lambda(x))]$ and C does not depend on λ, λ_0, x, or φ.

Lemma 1 shows that the conditional expectations are continuous in the parameter λ in the conditioning variable. This result is similar to Lemma A2(ii) of Song (2008). (See also Lemma A5 of Song (2009).)

We introduce an additional random vector $Z \in \mathbf{R}^{d_Z}$ with a support $\mathcal{S}_Z$ and a class Ψ being a class of $\mathbf{R}^J$-valued functions on $\mathbf{R}^{d_Z}$ with a generic element denoted by ψ and its envelope by $\tilde{\psi}$. As before, we fix $\lambda_0 \in \Lambda$, let $h_\lambda(z|\bar{\lambda}_1,\bar{\lambda}_2)$ denote the conditional density function of Z given $(\lambda_0(X),\lambda(X)) = (\bar{\lambda}_1,\bar{\lambda}_2)$ with respect to a σ-finite measure, and define $\mathcal{P}_Z \equiv \{h_\lambda(z|\cdot,\cdot) : (\lambda,z) \in \Lambda \times \mathcal{S}_Z\}$. Suppose that the parameter of interest takes the form of

$$\Gamma_{\varphi,\psi}(\lambda) = \mathbf{E}\left[\mu_\varphi(X;\lambda)^\top \psi(Z)\right].$$

We would like to analyze continuity of $\Gamma_{\varphi,\psi}(\lambda)$ in $\lambda \in \Lambda$. When $\mathcal{P}_Y$ and $\mathcal{P}_Z$ are regular, we obtain the following unexpected result.

Lemma 2 : *Suppose that $\mathcal{P}_Y$ is regular for $\tilde{\varphi}$ and $\mathcal{P}_Z$ is regular for $\tilde{\psi}$. Then, there exists $C > 0$ such that for each λ in Λ,*

$$\sup_{(\varphi,\psi)\in\Phi\times\Psi}|\Gamma_{\varphi,\psi}(\lambda) - \Gamma_{\varphi,\psi}(\lambda_0)| \leq C||\lambda - \lambda_0||_\infty^2.$$

Therefore, the first order Fréchet derivative of $\Gamma_{\varphi,\psi}(\lambda)$ at $\lambda_0 \in \Lambda$ is equal to zero.

Lemma 2 says that the functional $\Gamma_{\varphi,\psi}(\lambda)$ is not sensitive to the first order perturbation of λ around λ_0. In view of Newey (1994), Lemma 2 suggests that in general, there is no estimation effect of $\hat{\lambda}$ on the asymptotic variance of the estimator $\hat{\Gamma}_{\varphi,\psi}(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^n \hat{\mu}_\varphi(X_i;\hat{\lambda})^\top \psi(Z_i)$,