Journal of Econometrics 178 (2014) 471–483

Contents lists available at ScienceDirect

Journal of Econometrics

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

Semiparametric models with single-index nuisance parameters

Kyungchul Song

Vancouver School of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, BC, V6T 1Z1, Canada

ARTICLE INFO

Article history:
Received 16 October 2009
Received in revised form
24 June 2013
Accepted 4 July 2013
Available online 31 July 2013

JEL classification:
C12
C14
C51

Keywords:
Sample selection model
Conditional median restrictions
Matching estimators
Maximum score estimation
Cube-root asymptotics
Generated regressors

ABSTRACT

In many semiparametric models, the parameter of interest is identified through conditional expectations, where the conditioning variable involves a single-index that is estimated in the first step. Among the examples are sample selection models and propensity score matching estimators. When the first-step estimator follows cube-root asymptotics, no method of analyzing the asymptotic variance of the second step estimator exists in the literature. This paper provides nontrivial sufficient conditions under which the asymptotic variance is not affected by the first step single-index estimator regardless of whether it is root-*n* or cube-root consistent. The finding opens a way to simple inference procedures in these models. Results from Monte Carlo simulations show that the procedures perform well in finite samples.

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 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.[1]

Most literature deal with a single-index model as an isolated object, whereas empirical researchers often need to use the single-index specification in the context of estimating a larger model. An example is a structural model in labor economics that requires a prior estimation of components such as wage equations. When single-index components are nuisance parameters that are plugged into the second step estimation of a finite dimensional parameter of interest, the introduction of single-index restrictions does not improve the convergence rate of the estimated parameter of interest which already achieves the parametric rate of $\sqrt{n}$. Nevertheless, the use of a single-index restriction in such a situation still has its own merits. After its adoption, the model requires weaker assumptions on the nonparametric function and on the kernel function. This merit becomes prominent when the nonparametric function is defined on a space of a large dimension and stronger conditions on the nonparametric function and higher-order kernels are required. (See Hristache et al., 2001, for more details.)

This paper focuses on semiparametric models, where the parameter of interest is identified through a conditional expectation function and the conditioning variable involves a single-index with an unknown finite dimensional nuisance parameter. We assume that there is a consistent first step estimator of this nuisance parameter. In this situation, a natural procedure is a two step estimation, where one estimates the single-index first, and uses it to estimate the parameter of interest in the second step. Among the examples are sample selection models and propensity score matching estimators. The examples will be discussed in detail later.

A distinctive feature of the framework of this paper is that the first step estimator of a single-index is allowed to be either $\sqrt{n}$-consistent or $\sqrt[3]{n}$-consistent. The latter case of $\sqrt[3]{n}$-consistent single-index estimators is particularly interesting, for the framework includes new models that have not been studied in the literature, such as the sample selection model with conditional median restrictions, or propensity score matching estimators with conditional median restrictions. These conditional median restrictions

E-mail address: kysong@mail.ubc.ca.

[1] For example, Klein and Spady (1993) and Ichimura (1993) proposed *M*-estimation approaches to estimate the single-index, and Stoker (1986) and Powell et al. (1989) proposed estimation based on average derivatives. See also Härdle et al. (1993), Härdle and Tsybacov (1993), Horowitz and Härdle (1996), Fan and Li (1996) and Hristache et al. (2001).

0304-4076/$ – see front matter
http://dx.doi.org/10.1016/j.jeconom.2013.07.004

often lead to a substantial relaxation of the existing assumptions that have been used in the literature.[2]

Dealing with the case of a nuisance parameter that follows cube-root asymptotics of Kim and Pollard (1990) in two step estimation is challenging. In typical two step estimation, the asymptotic variance of the second step estimator involves an additional term due to the first step estimation of the single-index component (e.g. Newey and McFadden, 1994). Unless this term is shown to be negligible, one needs to compute this additional term by first finding the asymptotic linear representation of the first step estimator. However, in the case of a first step estimator that follows cube-root asymptotics, there does not exist such an asymptotic linear representation. This is precisely the situation where the first step estimator is a maximum score estimator proposed by Manski (1975).

The main contribution of this paper is to provide a set of conditions under which the first step estimator, regardless of whether it is $\sqrt{n}$-consistent or $\sqrt[3]{n}$-consistent, does not have an impact on the asymptotic variance of the second step estimator. This result is convenient, because under these conditions, one can simply compute the asymptotic variance as if one knows the true nuisance parameter in the single-index.

The result of this paper is based on a recent finding by the author (Song, 2012) which offers generic conditions under which conditional expectation functionals are very smooth. This smoothness is translated in our situation into insensitivity of the parameter of interest at a local perturbation of the single-index nuisance parameter.

To illustrate the usefulness of the result, this paper applies it to new semiparametric models such as semiparametric sample selection models with conditional median restrictions, and single-index matching estimators with conditional median restrictions. This paper offers procedures to obtain estimators and asymptotic variance formulas for the estimators.

This paper presents and discusses results from Monte Carlo simulation studies. The main focus of these studies lies on whether the asymptotic negligibility of the first step estimator's impact remains in force in finite samples. For this, it is investigated whether the estimators and the confidence sets based on the proposed asymptotic covariance matrix formula performs reasonably well in finite samples. Simulation results demonstrate clearly that they do so.

The main result of this paper is closely related to the literature of so-called *generated regressors* in nonparametric or semiparametric models. For example, Newey et al. (1999) and Das et al. (2003) considered nonparametric estimation of simultaneous equation models. Li and Wooldrige (2002) analyzed partial linear models with generated regressors when the estimated parameters in the generated regressors are $\sqrt{n}$-consistent. Rilstone (1996) and Sperlich (2009) studied nonparametric function estimators that involve generated regressors. Recent contributions by Hahn and Ridder (2010) and Mammen et al. (2012) offer a general analysis of the issue with generated regressors in nonparametric or semiparametric models. None of these papers considered generated regressors with coefficient estimators that follow cube-root asymptotics.

The paper is organized as follows. The paper defines the scope, introduces examples, and explains the main idea of this paper in Section 2. Then Section 3 presents the formal result of the asymptotic distribution theory, and discusses their implications for exemplar models. Section 4 discusses Monte Carlo simulation results, and Section 5 presents an empirical illustration based on a simple female labor supply model. Some technical proofs are found in the Appendix.

2. The scope, examples, and the main idea

2.1. The scope of the paper

Let us define the scope of the paper. Suppose that $W \equiv (W_1, \ldots, W_L)^\top \in \mathbf{R}^L$, S is a $d_S \times d_\varphi$ random matrix, and $X \in \mathbf{R}^d$ is a random vector, where all three random quantities W, S, and X, are assumed to be observable. We let $X = [X_1^\top, X_2^\top]^\top \in \mathbf{R}^{d_1+d_2}$, where X_1 is a continuous random vector and X_2 is a discrete random vector taking values from $\{x_1, \ldots, x_M\}$. Let $\Theta \subset \mathbf{R}^d$ be the space of a nuisance parameter θ_0 that is known to be identified. Denote $U_\theta \equiv F_\theta(X^\top\theta)$, where F_θ is the CDF of $X^\top\theta$. We assume that $X^\top\theta$ is a continuous random variable for all θ in a neighborhood of θ_0. Given an observed binary variable $D \in \{0, 1\}$, we define

$$\mu_\theta(U_\theta) \equiv \mathbf{E}\left[W|U_\theta, D = 1\right], \tag{1}$$

and when $\theta = \theta_0$, we simply write $\mu_0(U_0)$, where $U_0 \equiv F_{\theta_0}(X^\top\theta_0)$. The support of a random vector is defined to be the smallest closed set in which the random vector takes values with probability one. For $m = 1, \ldots, M$, let $\mathcal{S}_m$ be the support of $X1\{X_2 = x_m, D = 1\}$. And $\mathcal{S}_W$ be the support of W, and let $\varphi : \mathcal{S}_W \to \mathbf{R}^{d_\varphi}$ be a known map that is twice continuously differentiable with bounded derivatives on the interior of the support of $\mathbf{E}[W|X, D = 1]$. Then we define a map $a : \Theta \to \mathbf{R}^{d_S}$ by

$$a(\theta) \equiv \mathbf{E}\left[S \cdot \varphi(\mu_\theta(U_\theta))|D = 1\right], \quad \theta \in \Theta. \tag{2}$$

The general formulation admits the case without conditioning on $D = 1$ in which case it suffices to put $D = 1$ everywhere.[3]

This paper focuses on semiparametric models where the parameter of interest, denoted by β_0, is identified as follows:

$$\beta_0 = H(a(\theta_0), b_0), \tag{3}$$

where $H : \mathbf{R}^{d_S} \times \mathbf{R}^{d_b} \to \mathbf{R}^{d_\beta}$ is a map that is fully known, continuously differentiable in the first argument, and b_0 is a d_b dimensional parameter that does not depend on θ_0 and is consistently estimable. We will see examples of β_0 shortly.

Throughout this paper, we assume that there is an estimator $\hat{\theta}$ for θ_0 which is either $\sqrt{n}$-consistent or $\sqrt[3]{n}$-consistent. A natural estimator of β_0 is obtained by

$$\hat{\beta} \equiv H(\hat{a}(\hat{\theta}), \hat{b}),$$

where $\hat{a}(\theta)$ is an estimator of $a(\theta)$ and $\hat{b}$ is a consistent estimator of b_0. The estimator $\hat{a}(\theta)$ can be obtained by using nonparametric estimation of conditional expectation $\mathbf{E}\left[W|U_\theta, D = 1\right]$. For future reference, we denote

$$\tilde{\beta} \equiv H(\hat{a}(\theta_0), \hat{b}),$$

an infeasible estimator using θ_0 in place of $\hat{\theta}$. When $\hat{\theta}$ is $\sqrt[3]{n}$-consistent, it is not clear whether $\sqrt{n}(\hat{\beta}-\beta)$ will be asymptotically normal. In fact, it is not even clear whether $\hat{\beta}$ will be $\sqrt{n}$-consistent.

The main contribution of this paper is to provide conditions under which, whenever $\hat{\theta} = \theta_0 + O_P(n^{-1/3})$ and

$$\sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{d} N(0, V), \tag{4}$$

it follows that

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V).$$

This result is very convenient, because the computation of the asymptotic variance matrix V in (4) can be done, following the standard procedure.

---

[2] For example, the semiparametric sample selection model in Newey et al. (1990) assumes that the error term in the selection equation is independent of observed covariates. Also, parametric specifications of propensity scores in the literature of program evaluations (such as logit or probit specifications) assume that the error term in the program participation equation is independent of observed covariates. (See Heckman et al., 1998a, for example.). In these situations, the assumption of the conditional median restriction is a weaker assumption because it allows for stochastic dependence between the error term and the observed covariates.

[3] The conditions for the identification of θ_0 in many examples of semiparametric models is already established. See Horowitz (2009). The identification of θ_0 in this paper's context arises often in binary choice models. See Chapter 4 of Horowitz (2009) for the identification analysis for the binary choice models.