Efficient estimation of partially linear varying coefficient models

Wei Long\textsuperscript{a}, Min Ouyang\textsuperscript{b}, Ying Shang\textsuperscript{c,}\textsuperscript{*}

\textsuperscript{a} Department of Economics, Texas A\&M University, USA
\textsuperscript{b} Department of Economics, Tsinghua SEM, Beijing, PR China
\textsuperscript{c} International School of Economics & Management, Capital University of Economics & Business, Beijing, PR China

\textbf{HIGHLIGHTS}

\begin{itemize}
  \item We derive the semiparametric efficient bound for a partially linear varying coefficient model.
  \item We show that a kernel estimator is semiparametrically efficient.
  \item Simulation results strongly support out theoretical analysis.
\end{itemize}

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 15 May 2013
Received in revised form 10 July 2013
Accepted 13 July 2013
Available online 18 July 2013

\textbf{JEL classification:}
C14

\textbf{Keywords:}
Partially linear
Varying coefficient
Semiparametric method
Efficient estimation
Simulation

\textbf{ABSTRACT}

In this paper, we consider the problem of estimating a semiparametric partially linear varying coefficient model. We derive the semiparametric efficiency bound for the asymptotic variance of the finite-dimensional parameter estimator. We also propose an efficient estimator for estimating the finite-dimensional parameter of the model. Simulation results show substantial efficiency gain of our proposed estimator over a conventional estimator as considered in Ahmad et al. (2005).

© 2013 Elsevier B.V. All rights reserved.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Graphical representation of the model.}
\end{figure}

1. Introduction

Semiparametric regression models have the advantage of providing flexible regression functional form, and at the same time they avoid the 'curse of dimensionality' of fully nonparametric regression models. Partially linear models (e.g., Robinson, 1988), varying coefficient models (Cai et al., 2000), and additive models (Linton and Nielsen, 1995) are well-studied semiparametric models. These semiparametric models are also among the popular specifications adopted by applied researchers. For applications of these models in economics, see, for example, Li et al. (2002) and Stengos and Zacharias (2006).

Efficient estimation of the finite-dimensional parameters in a semiparametric model is an important research topic. Chamberlain (1992) studied the problem of efficient estimation of a semiparametric partially linear model. Ma et al. (2006) provided extensive simulations on six different estimators of a partially linear model and found that the estimator with the optimal weight function indeed performs the best. In this paper, we examine the problem of efficient estimation of a partially linear varying coefficient model. We derive the semiparametric efficient bound for the linear parameter of the model. Replacing population quantities by sample analogues, we obtain a feasible estimator that mimics the infeasible semiparametric efficient estimator. Simulations are used to examine the finite-sample performance of our feasible estimator.

2. The model and the estimator

We consider the following partially linear varying coefficient model:

\[
y_i = X_i'\gamma + w_i'\beta(z_i) + u_i, \quad i = 1, \ldots, n
\] (2.1)
where $x_i$ and $w_i$ are vectors of dimensions $d$ and $q$, $\gamma$ is a $d \times 1$ vector of constant parameter, $\beta(z)$ is a $q \times 1$ vector of unspecified functions, and $y_i$, $z_i$, and $u_i$ are scalars.

Ahmad, Leelahanon and Li (2005, ALL) considered the estimation of the above partially linear varying coefficient (PL-VC) model using the series estimation method. They show that their estimator of $\gamma$ is semiparametrically efficient when the error $u_i$ is conditional homoskedastic. In this paper we generalize ALL’s result by proposing a kernel-based estimator of $\gamma$ that is efficient when the error $u_i$ is conditional heteroskedastic.

Let $v_i = (w_i', x_i', z_i)$ and $\sigma^2(v_i) = E(u_i^2 | v_i)$. Following Li and Racine (2006, pp. 235–236), we estimate $\gamma$ and $\beta(z_i)$ by minimizing the following objective function:

$$\inf_{\gamma \in \Gamma, \beta \in \mathcal{F}} E \left\{ \left[ y_i - x_i' \gamma - w_i' \beta(z_i) \right]^2 / \sigma^2(v_i) \right\},$$

(2.2)

where $\Gamma$ is a compact subset of $\mathbb{R}^d$ and $\mathcal{F}$ is a class of smooth functions as defined in ALL (2005).

In an application, we need to replace the population mean by the sample mean, and we choose $\hat{\gamma}$ and $\hat{\beta}(z)$ to minimize

$$\min_{\gamma \in \Gamma, \beta \in \mathcal{F}} \sum_{i=1}^n \left[ y_i - x_i' \gamma - w_i' \beta(z_i) \right]^2 / \sigma^2(v_i).$$

(2.3)

We first treat $\gamma$ as known; then, by the application of calculus of variation to $\beta(z_i)$ in (2.2), we get

$$-2 E \left[ w_i' (y_i - x_i' \gamma - w_i' \beta(z_i)) / \sigma^2(v_i) \right] \delta(z_i) = 0,$$

(2.4)

where $\delta(z_i)$ is an arbitrary function of $z_i$. Eq. (2.4) implies that $E \left[ w_i (y_i - x_i' \gamma - w_i' \beta(z_i)) / \sigma^2(v_i) | z_i \right] = 0$. Solving for $\beta(z)$ gives

$$\beta(z_i) = \left[ E \left( \frac{w_i w_i'}{\sigma^2(v_i)} | z_i \right) \right]^{-1} E \left( \frac{w_i (y_i - x_i' \gamma)}{\sigma^2(v_i)} | z_i \right)$$

(2.5)

$$= A_{1i} - A_{2i} \gamma,$$

where $A_{1i} = \left[ E \left( \frac{w_i w_i'}{\sigma^2(v_i)} | z_i \right) \right]^{-1} E \left( \frac{w_i y_i}{\sigma^2(v_i)} | z_i \right)$ and

$$A_{2i} = \left[ E \left( \frac{w_i w_i'}{\sigma^2(v_i)} | z_i \right) \right]^{-1} E \left( \frac{w_i x_i}{\sigma^2(v_i)} | z_i \right).$$

Substituting (2.5) into (2.3), we obtain

$$\min_{\gamma \in \Gamma} \sum_{i=1}^n \left[ y_i - w_i' A_{1i} - (x_i' - w_i' A_{2i}) \gamma \right]^2 / \sigma^2(v_i)$$

$$= \sum_{i=1}^n \left[ \tilde{y}_i - \tilde{x}_i \gamma \right]^2 / \sigma^2(v_i),$$

(2.6)

where $\tilde{y}_i = y_i - w_i' A_{1i}$, $\tilde{x}_i = x_i - A_{2i} w_i$, and $A_{1i}$ and $A_{2i}$ are defined below (2.5).

Minimizing (2.6) with respect to $\gamma$ gives

$$\hat{\gamma} = \left[ \sum_{i=1}^n \tilde{x}_i \tilde{x}_i / \sigma^2(v_i) \right]^{-1} \left[ \sum_{i=1}^n \tilde{x}_i \tilde{y}_i / \sigma^2(v_i) \right].$$

(2.7)

From (2.7), we immediately have

$$\sqrt{n} (\hat{\gamma}_{\inf} - \gamma) \xrightarrow{d} N(0, \Sigma^{-1}),$$

(2.8)

where $\Sigma = E[\tilde{x}_i \tilde{x}_i / \sigma^2(v_i)]$ with $\tilde{x}_i' = x_i' - w_i' E \left( \frac{w_i w_i'}{\sigma^2(v_i)} | z_i \right)^{-1} E \left( \frac{w_i y_i}{\sigma^2(v_i)} | z_i \right)$. It can be shown that $\Sigma$ is the semiparametric variance lower bound for any regular estimator of $\gamma$.

However, $\hat{\gamma}_{\inf}$ defined in (2.7) is not feasible, because it involves unknown conditional mean functions as well as unknown conditional variance function $\sigma^2(v_i)$. In order to obtain a feasible efficient estimator, we need to estimate $\sigma^2(v_i)$ and $\gamma$ by $E(u_i^2 | v_i)$ and $E(u_i^2 | v_i)$. This can be done by first replacing $\sigma^2(v_i)$ by 1, and replacing the conditional mean functions $E(B_i | z_i)$ by kernel estimator $\hat{E}(B_i | z_i) = \sum_i B_i K_{x_i / h} / \sum_i K_{x_i / h} = K_i / h_i$ (2.7), where $B_i = w_i x_i / h_i$ or $B_i = w_i x_i / h_i = K_i / h_i$ is the kernel function, and $h_i$ is the smoothing parameter. We obtain a feasible estimator of $\gamma$ as follows:

$$\hat{\gamma} = \left[ \sum_{i=1}^n \tilde{x}_i \tilde{x}_i / \sigma^2(v_i) \right]^{-1} \left[ \sum_{i=1}^n \tilde{x}_i \tilde{y}_i / \sigma^2(v_i) \right].$$

(2.9)

Substituting $\hat{\beta}(z_i) = \hat{E}(w_i | y_i) / \sigma^2(v_i)$ into (2.9), respectively, gives a feasible estimator of $\beta(z_i)$:

$$\hat{\beta}(z_i) = \left[ \hat{E}(w_i | y_i) / \sigma^2(v_i) \right]^{-1} \left[ \hat{E}(w_i | y_i - x_i' \hat{\gamma}) / \sigma^2(v_i) \right]$$

$$= \hat{y}_i - \tilde{x}_i \hat{\gamma},$$

where the definitions of $\hat{y}_i$ and $\tilde{x}_i$ should be apparent.

With $\hat{\gamma}$ and $\hat{\beta}(z)$ defined in (2.9) and (2.10), we get an estimator of $\theta_i$ by $\hat{u}_i = y_i - \hat{\gamma} - \hat{\beta}(z_i)$. Then we can estimate $\sigma^2(v_i)$ by $\hat{\sigma}_i^2 = \sum_i \hat{u}_i^2 / n$, where $\hat{u}_i = K_{x_i / h} / \sum_i K_{x_i / h}$ and $K_{x_i / h} = \hat{K}(x_i / h)$. Hence, $\sigma^2(v_i)$ and $\hat{\gamma}$ are kernel functions associated with $x_i$ and $w_i$, $h_i$ are the corresponding smoothing parameters. When the dimension of $x_i$ is greater than one, the kernel is the product kernel function and $h_i$ is a vector of smoothing parameters. The semiparametric efficient estimator of $\gamma$ is given by

$$\sqrt{n} (\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \Sigma^{-1}),$$

where $\Sigma = E[\tilde{x}_i \tilde{x}_i / \sigma^2(v_i)]$ with $\tilde{x}_i' = x_i' - u_i' E \left( \frac{w_i w_i'}{\sigma^2(v_i)} | z_i \right)^{-1} E \left( \frac{w_i y_i}{\sigma^2(v_i)} | z_i \right)$. It can be shown that $\Sigma$ is the semiparametric variance lower bound for any regular estimator of $\gamma$.

A sketchy proof of Theorem 1 is given in the Appendix.

3. Simulation results

In this section, we examine the finite-sample performance of $\hat{\gamma}$ and compare it with that of $\hat{\gamma}$. We generate $x_i$ i.i.d. uniform[0, 2], $w_i$ i.i.d. uniform[0, 2], and $z_i$ i.i.d. uniform[0, 1], $u_i = \epsilon_i \sigma(v_i)$, $\sigma^2(v_i) = 0.5 \sigma^2(x_i + w_i z_i)$, and $\epsilon_i$ is i.i.d. $N(0, 0.25)$. The dependent variable $y_i$ is generated by

$$y_i = y_0 + x_i \gamma_1 + w_i \beta(z_i) + u_i,$$

with $y_0 = 1$, $\gamma_1 = 1$, and $\beta(z_i) = \sin(2 \pi z_i)$. The sample sizes are $n = 100, 200, 400,$ and $800$. The number of replications is 5000. We examine the estimated mean squared error by $\text{MSE}(\hat{\gamma}) = M^{-1} \sum_{s=1}^M (\hat{\gamma}_s - \gamma_1)^2$, for $s = 0, 1$, where $M$ is the number of replications.
replications. For comparison purposes, we also report the estimated MSE of \( \hat{y} \). The estimation results are reported in Table 1.

From Table 1, we first observe that, when the sample size is doubled, the MSE of \( \hat{y} \) and \( \tilde{y} \) is reduced by half; this is as expected, because both \( \hat{y} \) and \( \tilde{y} \) are \( \sqrt{n} \)-consistent estimators for \( y \). The MSE of \( \hat{\beta}(\cdot) \) and \( \hat{\beta}(\cdot) \) also decreases as \( n \) increases, suggesting consistency of these estimators. Second, from Table 1, we can see that the MSE of \( \tilde{y} \) is about half of the MSE of \( \hat{y} \). This confirms our theoretical result that \( \tilde{y} \) is more efficient than \( \hat{y} \), because the former is semiparametrically efficient while the latter is not. For the estimation of the nonparametric component \( \beta(\cdot) \), we note that \( \hat{\beta}(\cdot) \) is only slightly more efficient than \( \hat{\beta}(\cdot) \). This is also as expected, because both \( \hat{y} \) and \( \tilde{y} \) are \( \sqrt{n} \)-consistent estimators of \( y \). Also, since \( O_p(n^{-1/2}) \) is smaller than the nonparametric estimation error rate, \( \hat{\beta}(\cdot) \) and \( \hat{\beta}(\cdot) \) are asymptotically equivalent to each other, and both are asymptotically equivalent to an infeasible estimator of \( \hat{\beta}(\cdot) \) that uses the true value of \( y \). In sum, the simulation results reported in this section strongly support our theoretical analysis in this paper.

Acknowledgment

We would like to thank Professor Qi Li for his helpful guidance throughout this research.

Appendix. A sketchy proof of Theorem 1

The key step for proving Theorem 1 is to use the uniform convergence rate of nonparametric estimator \( \hat{\sigma}^2(v) = \sigma^2(v) \). To simplify the exposition, we will only consider the case that \( x_i \) and \( w_i \) are scalars. Instead of giving the lengthy primitive regularity conditions, we simply assume that the uniform rate holds for \( \hat{\sigma}^2(v) = \sigma^2(v) \), and discuss what kind of conditions are needed to ensure the uniform convergence. We consider i.i.d. data, and let \( V \) denote the support of \( v_i = (x_i, w_i, z_i) \); then \( V \) is a compact subset of \( \mathbb{R}^3 \). The density function \( f(v) \) is bounded below and above by two positive constants.

A trimming set \( M_0 \) is also needed to trim data near the boundary of the support. Following Li et al. (2013), we trim sufficiently small intervals off the boundaries when the boundaries of the support \( V \) are known. However, in practice the boundaries of \( V \) are usually unknown. In this case, we sort the data to find the sample minimums and maximums of \( v_i \). With sufficiently large sample size, such sample minimums and maximums converge to the true boundaries of \( V \). Hence, we just trim sufficiently small intervals of the sample boundaries when the true boundaries of \( V \) are unknown. Specifically, let \( V = [a_w, b_w] \times [a_x, b_x] \times [a_z, b_z] \) denote the support of \( v_i = (x_i, w_i, z_i) \). When \( a_w \) and \( b_w \) are unknown (\( \theta = (x, w, z) \)), we use \( a_w = \min_{1 \leq i \leq n} a_{w_i} \) and \( b_w = \max_{1 \leq i \leq n} b_{w_i} \) to estimate \( a_w \) and \( b_w \), respectively. It is well established that \( a_w = a_w \to 0 = O_p(n^{-1}) \) and \( b_w = b_w = O_p(n^{-1}) \) (they are super-consistent estimators). The trimming set can be chosen as \( M_0 = [\min_{1 \leq i \leq n} a_{w_i} + \delta_w, \max_{1 \leq i \leq n} b_{w_i} - \delta_w] \times [\min_{1 \leq i \leq n} a_{x_i} + \delta_x, \max_{1 \leq i \leq n} b_{x_i} - \delta_x] \times [\min_{1 \leq i \leq n} a_{z_i} + \delta_z, \max_{1 \leq i \leq n} b_{z_i} - \delta_z] \), where \( \delta_w = \delta_w = \delta_x = 0 = \delta_z \to 0 \) as \( n \to \infty \), and \( h_y/h_0 = o(1), \theta = (x, w, z) \).

We assume that
\[
\sup_{v \in M_0} |\hat{\sigma}^2(v) - \sigma^2(v)| = O_p \left( h_x^2 + h_w^2 + h_z^2 + \left\{ \frac{\ln(n)}{n} h_w w \right\}^{1/2} \right). \tag{A.1}
\]
Since \( \hat{\sigma}^2(v) \) uses \( \{\hat{a}_w \}^n_{i=1} \), (A.1) requires, among other things, the uniform convergence rate \( \sup_{v \in M_0} |\hat{\beta}(z) - \beta(z)| = O_p \left( h_x^2 + \left\{ \frac{\ln(n)}{n} h_w w \right\}^{1/2} \right) \), where \( M_{n,z} = [\hat{a}_x + \delta_x, \hat{b}_x - \delta_x] \) is the trimmed subset of the support of \( z \). The proof of Theorem 1 uses (A.1) and the following identity:
\[
\frac{1}{\sigma^2(v)} = \frac{1}{\hat{\sigma}^2(v)} + \frac{\sigma^2(v) - \hat{\sigma}^2(v)}{\sigma^2(v) \hat{\sigma}^2(v)} = \frac{1}{\sigma^2(v)} + \frac{\sigma^2(v) - \hat{\sigma}^2(v)}{\sigma^4(v) \hat{\sigma}^2(v)} + \frac{(\sigma^2(v) - \hat{\sigma}^2(v))^2}{\sigma^4(v) \hat{\sigma}^2(v) \sigma^2(v)}, \tag{A.2}
\]
Now, define another infeasible estimator \( \hat{y}_{\text{inf}} \), which can be obtained from \( \tilde{y} \) as given in (2.11), except that \( \hat{\sigma}^2(v) \) in \( \tilde{y} \) is replaced by \( \sigma^2(v) \). Substituting (A.2) into the definition of \( \tilde{y} \), then, under some standard conditions (assuming that \( h = h_u = h_w = h_z \)), including \( nh_w \to 0 \) and \( nh_w/\ln(n) \to 0 \) as \( n \to \infty \), it can be shown that the leading term of \( \tilde{y} \) is given by \( \hat{y}_{\text{inf}} \). Under some standard regularity conditions, including smoothness conditions on \( \beta(z), f(v) \) and some conditions on \( h \) such as those given in Fan and Huang (2005), it is straightforward to show that \( \sqrt{n} (\hat{y}_{\text{inf}} - y) \to N(0, \Sigma^{-1}) \). Then Theorem 1 follows from \( \sqrt{n} (\tilde{y} - y) = \sqrt{n} (\hat{y}_{\text{inf}} - y) + o_p(1) \).

References


