Generalized varying coefficient models with unknown link function

BY C. N. KURUWITA, K. B. KULASEKERA AND C. M. GALLAGHER

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina 29634, U.S.A.
chinthk@g.clemson.edu kk@clemson.edu cgallag@clemson.edu

SUMMARY

We propose a new estimation method for generalized varying coefficient models where the link function is specified up to some smoothness conditions. Consistency and asymptotic normality of the estimated varying coefficient functions are established. Simulation results and a real data application demonstrate the usefulness of the new method.

Some key words: Single index model; Varying coefficient model.

I. Introduction

The varying coefficient model has gained considerable interest since pioneering work by Hastie & Tibshirani (1993). Its ability to model dynamical systems led to applications in areas including functional data modelling (Ramsay & Silverman, 1998), time series analysis (Huang & Shen, 2004), longitudinal data analysis (Fan et al., 2007), survival analysis (Cai et al., 2008) and nonparametric quantile regression (Cai & Xu, 2009).

Suppose we have independent and identically distributed observations $(Y_i, X_i, U_i)$ following the generalized varying coefficient model

$$Y_i = g \left\{ \sum_{k=1}^{p} \beta_k(U_i)X_{ik} \right\} + \epsilon_i \tag{1}$$

with $E(\epsilon_i \mid U_i, X_i) = 0$ $(i = 1, \ldots, n)$. Here $g(\cdot)$ is called the link function and the $\beta_k(\cdot)$s are the varying coefficient functions. The covariate $X_i$ is $p$-dimensional and $U_i$ is a univariate random variable called the effect modifier or index variable. Cai et al. (2000) studied the estimation and hypothesis testing of the varying coefficient functions in model (1) with a known link function. However, assuming a parametric form for the link function is very restrictive and its misspecification can result in large bias in the estimated varying coefficient functions. Therefore, it is desirable to have the link function unspecified in model (1), especially at the exploratory stage of modelling. This note discusses nonparametric estimation of the varying coefficient functions in model (1) when the link function is unknown.

Regression models with unknown link functions have been studied by several authors in the context of generalized linear models and generalized additive models. Li & Duan (1989) discuss asymptotic properties of the estimated regression coefficients under link violation and give conditions needed to attain consistency of the estimated regression parameters. Weisberg & Welsh (1994) proposed to estimate the unknown link function using a Nadaraya–Watson kernel
smoother and used an iterative weighted least squares method to estimate the regression coefficients. Extending this idea, Chiu & Müller (1998) proposed a quasilikelihood approach and established the asymptotic normality of the regression parameters with unknown link and variance functions. Horowitz (2001) studied the estimation in a generalized additive model with an unknown link function and showed that the additive components and the link function can be estimated consistently. However, no work has been done to extend these methods to varying coefficient models with unknown links.

We introduce an estimation method that can be used to estimate the varying coefficient functions of model (1) with an unspecified link function. Our approach involves a simple localized least-squares minimization which is noniterative in the estimate/update/re-estimate steps. It not only gives consistent estimates of the coefficient functions but also allows us to estimate the unknown link, which can suggest a parametric link function for the data.

2. Estimation method and asymptotic properties

2.1. Estimation method

For convenience, assume the coefficient functions \( \beta_k(\cdot) \) \((k = 1, \ldots, p)\) of model (1) are defined on \([0, 1]\) with each \( \beta_k(\cdot) \) having a continuous derivative. Our aim is to estimate these coefficient functions pointwise. Therefore, let \( \eta_i(u_0) \) be the local constant approximation of the linear predictor \( \eta(U_i) = \sum_{k=1}^{p} \beta_k(U_i) X_{i,k} \) of model (1), for \( 0 \leq u_0 \leq 1 \). Then \( \eta_i(u_0) = \theta^T X_i \), where \( \theta = (a_1, \ldots, a_p)^T \) and \( X_i = (X_{i1}, \ldots, X_{ip})^T \). If the link function \( g \) was known, one could obtain local estimates of the coefficients by a straightforward weighted least squares minimization. Since \( g \) is unspecified in our model, a natural strategy is to estimate the link function nonparametrically. However, since the coefficient functions are unknown, the linear predictor is unknown, so we cannot simply smooth the responses against the linear predictor to estimate the link function. Noting that \( \eta_i(u_0) = \theta^T X_i \) is a local approximation of the linear predictor, an estimate of the link function can instead be obtained by smoothing \( \{\theta^T X_i, Y_i\} \) \((i = 1, \ldots, n)\). For example, let \( \hat{\eta}_{NW}(t, \theta) = \sum_{i=1}^{n} w_i Y_i \) be the Nadaraya–Watson estimator of the link function at \( t \), where \( w_i = K_h(\theta^T X_i - t) / \sum_{j=1}^{n} K_h(\theta^T X_j - t) \) with \( K_h(\cdot) = K(\cdot/h) \) is a symmetric kernel function. However, if we simply use a standard one-dimensional smoother of this form, we are ignoring the fact that \( \eta_i(u_0) \) is only a localized estimate of the true linear predictor for a given \( U = u_0 \), and as a result the link function estimate will exhibit very poor finite sample performance as shown in Fig. 1. Furthermore, such an estimate will be inconsistent. Suppose we know the true values \( \theta_0 = (\beta_1(u_0), \ldots, \beta_p(u_0))^T \) of the coefficient functions of model (1) at \( U = u_0 \). Under suitable conditions on the smoothing parameter \( h \), \( \hat{\eta}_{NW}(t, \theta_0) \to E(Y \mid \theta_0^T X = t) \) in probability as \( n \to \infty \). However, \( E(Y \mid \theta_0^T X = t) = g(t) = E(Y \mid \theta_0^T X = t, U = u_0) \). Therefore, in order to consistently estimate the link function, we need to ensure that only the observations close to \( u_0 \) are used in the smoothing process. To achieve this objective, we use two kernels in a Nadaraya–Watson type estimator to get a localized estimate of the link function.

For \( t \) on \( T \), the support of \( \theta^T X \), let \( \hat{\eta}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta) \) where

\[
A_{n,u_0}(t, \theta) = (nh_1 h_2)^{-1} \sum_{j=1}^{n} Y_j K_1 \left( \frac{\theta^T X_j - t}{h_1} \right) K_2 \left( \frac{U_j - u_0}{h_2} \right),
\]

and \( B_{n,u_0}(t, \theta) \) is \( A_{n,u_0}(t, \theta) \) with \( Y_j \equiv 1 \) \((j = 1, \ldots, n)\). The kernel \( K_1(\cdot) \) with smoothing parameter \( h_1 \) localizes the observations around the point of estimation \( t \) as in the Nadaraya–Watson estimator, while the kernel \( K_2(\cdot) \) with smoothing parameter \( h_2 \) localizes the observations...
around \( u_0 \). The estimator \( \hat{g}_{u_0}(t, \theta) \) is similar to that of Ichimura (1993). However, the weight function in his estimator does not depend on the sample size and is used as a means of handling heteroscedasticity. In contrast, the weights given by the extra kernel \( K_2(\cdot) \) in our estimator depend on the sample size and serve the purpose of localizing the observations around the point \( u_0 \).

Given this estimator of the link function, we use the following estimation procedure to estimate the varying coefficient functions.

**Step 1.** On a grid of \( u_0 \) values in \([0, 1]\), minimize the localized least squares criterion

\[
M_n(\theta) = (nh_2)^{-1} \sum_{i=1}^{n} (Y_i - \hat{g}_{u_0}(\theta^T X_i, \theta))^2 K_2 \left( \frac{U_i - u_0}{h_2} \right) \tag{3}
\]

to estimate \( \beta_k(\cdot) \) \((k = 1, \ldots, p)\). For each \( u_0 \) the minimizer \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p) \) of (3) is the estimate of \( \beta_k(u_0) \).

**Step 2.** Once the coefficient functions are recovered, an improved estimate of the link function \( \hat{g}(\cdot) \) is constructed in the following way. Noting that the model shares a common link function for each fixed \( u_0 \) let \( \hat{\eta}(U_i) = \sum_{k=1}^{p} \hat{\beta}_k(U_i) X_i \) be an estimate of the linear predictor of model (1). Smoothing \( \{\hat{\eta}(U_i), Y_i\} \) \((i = 1, \ldots, n)\) with a Nadaraya–Watson estimator we get

\[
\hat{g}(t) = \frac{\sum_{i=1}^{n} K_1((\hat{\eta}(U_i) - t)/h_1)Y_i}{\sum_{i=1}^{n} K_1((\hat{\eta}(U_i) - t)/h_1)}.
\]

**Remark 1.** In our estimation of the coefficient functions and the link function, we have used a local constant approximation. However, one could easily use local linear smoothing in estimating the coefficient functions and the link function. For example, a local linear approximation of the linear predictor can be written as \( \eta_i^{LOL}(u_0) = \gamma^T Z_i \) where \( \gamma = (a_1, \ldots, a_p, b_1, \ldots, b_p)^T \) and \( Z_i = (X_{i1}, \ldots, X_{ip}, (U_i - u_0)X_{i1}, \ldots, (U_i - u_0)X_{ip})^T \). Substituting \( \gamma \) for \( \theta \) and \( Z_i \) for \( X_i \) in (2) and (3) will yield local linear estimates of the coefficient functions.

Bandwidth selection is an important aspect of any nonparametric estimation problem. For our estimation procedure, we need to select two bandwidths \( h_1 \) and \( h_2 \). In our data example, we use...
leave-one-out crossvalidation to select them, using the crossvalidation function

$$CV(h_1, h_2) = \sum_{i=1}^{n} (Y_i - \hat{Y}^{-i})^2,$$

where $\hat{Y}^{-i}$ is the fitted value with the $i$th observation removed and using bandwidth combination $(h_1, h_2)$. We minimize this function over a two-dimensional grid of bandwidth values and choose the bandwidth combination which yields the minimum crossvalidation score. When the sample size is large computing the leave-one-out crossvalidated bandwidths is time-consuming. Therefore, in our simulations we used a 5-fold crossvalidation method.

2.2. Asymptotic properties

For a specified link function $g(\cdot)$, Cai et al. (2000) established pointwise asymptotic normality of the estimated varying coefficient functions in model (1). Extending their work, we show that our estimated varying coefficient functions with an unspecified link are consistent and asymptotically normal. We have the following results.

**Theorem 1.** Under Conditions A1–A8 in the appendix, the minimizer $\hat{\theta}$ of (3) is a consistent estimator of $\theta_0$.

**Theorem 2.** Under Conditions A1–A9 in the appendix, the minimizer $\hat{\theta}$ of (3) converges in distribution to a normal random variable with mean vector $\theta_0$ and covariance matrix

$$\Sigma_{u_0} = (M_2(\theta_0))^{-1} f_U(u_0) \sigma_0 \Delta(u_0), \quad \nu_j = \int s^j \kappa_2^2(s) ds.$$

**Remark 2.** The order of the bias in the coefficient function estimator is the same as that of Cai et al. (2000) and is given by (A6) in the appendix. Due to the restrictions on the bandwidth sequences, the asymptotic bias of our coefficient function estimator becomes zero at the expense of a relatively slower convergence rate than that of the coefficient function estimator with known link. In practice, one could use a method of moment estimator of $\Sigma_{u_0}$ similar to Cai et al. (2000) for inference. In our data example, we used bootstrap methods to estimate the standard errors of the coefficient function estimates.

3. Empirical study

3.1. Simulation results

We generated 1000 random samples of size $n$ responses from model (1) with three covariates and link functions $g(t) = t$, $g(t) = t^2$ and $g(t) = \sin(2t)$. The coefficient functions $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\beta_3(\cdot)$ were chosen to be the normalized versions of $t^2 + 1$, $\cos^2(\pi t) + 0.5$ and $2 \sin^2(\pi t) - 0.5$ that satisfy the identifiability Condition A3 in the appendix. We took $X_1$, $X_2$ and $X_3$ to be independent standard normal covariates, the effect modifier $U$ to be a uniform random variable over [0, 1] and $\epsilon$ to be a normal random variable with zero mean and standard deviation 0.1 independent of both $X$ and $U$. We used the standard Gaussian kernel for $K_1(\cdot)$ and the Epanechnikov kernel $K(s) = 0.75(1 - s^2)_+$ as $K_2(\cdot)$. The performance of the estimated
Table 1. Comparison of average squared error, $\text{ASE}_{\beta}$, of the coefficient function estimates from the proposed estimation method with the unknown link and the estimation method with the known link. Results given are the mean $\text{ASE}_{\beta}$ values from 1000 simulations with crossvalidated bandwidths.

<table>
<thead>
<tr>
<th>Sample size $n$</th>
<th>Known link</th>
<th>Unknown link</th>
<th>Known link</th>
<th>Unknown link</th>
<th>Known link</th>
<th>Unknown link</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.562</td>
<td>0.965</td>
<td>0.701</td>
<td>2.623</td>
<td>1.101</td>
<td>2.460</td>
</tr>
<tr>
<td>200</td>
<td>0.219</td>
<td>0.359</td>
<td>0.327</td>
<td>1.385</td>
<td>0.227</td>
<td>0.627</td>
</tr>
<tr>
<td>400</td>
<td>0.105</td>
<td>0.177</td>
<td>0.058</td>
<td>0.360</td>
<td>0.077</td>
<td>0.241</td>
</tr>
</tbody>
</table>

varying coefficient functions and the link function is assessed using average squared errors

$$\text{ASE}_{\beta} = N_{\beta}^{-1} \sum_{k=1}^{3} \sum_{j=1}^{N_{\beta}} (\hat{\beta}_k(u_j) - \beta_k(u_j))^2,$$

$$\text{ASE}_g = N_g^{-1} \sum_{j=1}^{N_g} (\hat{g}(t_j) - g(t_j))^2,$$

respectively. Here $0 \leq u_j \leq 1$ ($j = 1, \ldots, N_{\beta}$) and $-2 \leq t_j \leq 2$ ($j = 1, \ldots, N_g$) are gridpoints at which we estimate the coefficient functions and the link function. We used $N_{\beta} = 100$ and $N_g = 200$.

As shown in Fig. 1, the use of a one-dimensional smoother such as the Nadaraya–Watson estimator in estimating the link function is inefficient. The $\text{ASE}_g$ values, not presented here, confirm the need for additional localizing with respect to the index variable $U$. For example, 1000 simulations with a linear link and sample size of 200, resulted in a mean $\text{ASE}_g$ of 0.2 and 109.73 for our link function estimator and the one-dimensional Nadaraya–Watson estimator, respectively.

We compared the performance of our coefficient function estimator with that of Cai et al. (2000). Table 1 summarizes the mean $\text{ASE}_{\beta}$ for known link and unknown link methods. Our estimator tends to have larger average squared error values compared with the estimator with known link, due to the additional estimation step involved in our method. Both estimators show a decrease in average squared error values as sample size increases. In order to assess the pointwise variability of the coefficient function estimators, we plotted the 25th and the 75th pointwise percentiles of the 1000 estimates of the coefficient functions. To save space, we only present the case for $\beta_1(\cdot)$ with quadratic link in Fig. 2.

3.2. Real data example

To illustrate our methodology, we analyse the Japanese chemical industry data (Yafeh & Yosha, 2003) which is publicly available at http://www.res.org.uk. Data consist of various economic factors collected on 185 chemical firms listed in the Japanese stock market. The dependent variable $Y$ is a measure of expenses on managerial private benefits. Three covariates are considered: the age of the firm; ownership concentration, which gives the percentage of ownership that belongs to the top 10 shareholders; and profit of the firm. They are denoted by Age, Top10 and Profit, respectively. As our effect modifier $u$, we picked leverage, which is the ratio of debt to debt plus equity. This allows us to examine how the effects of the covariates on the response changes with the firm’s debt/equity levels. We used leave-one-out cross-validation to select the two smoothing parameters $h_1$ and $h_2$.

Figure 3(a)–(e) shows the estimated coefficient functions and the link function estimate together with 95% bootstrap confidence intervals based on 1000 bootstrap replications. All three confidence intervals of the coefficients exclude zero in most of the support which indicates that
Fig. 2. Pointwise 25th and 75th percentiles of the estimates of coefficient function $\beta_1(\cdot)$ with link function $g(t) = t^2$ and sample size $n = 200$. True function (solid), 25th percentile (dashed) and 75th percentile (dotted).

Fig. 3. Analysis of Japanese chemical industry data. (a)–(c) Estimated coefficient functions, (d) link function estimate, (e) cross-validation surface, (f) prediction errors for known link and unknown link methods.

all three covariates have a significant impact on the response. Figure 3(b) suggests that ownership concentration significantly affects the response in firms that have higher leverage. The link function estimate Fig. 3(d) appears to be monotonically increasing for most of the support but remains nonlinear.
The natural link function for these data is the linear link. In order to assess the performance of our model with that of the known link method, we compared the average prediction errors of the two methods by partitioning the data into a training set of 150 observations and a test set of 35 observations. We performed this for 10 random partitions of the data, with leave-one-out crossvalidated bandwidth selection on each of the 10 partitions for both the known link and unknown link methods. The average prediction errors summarized in Fig. 3(f) show that our method exhibits a lower prediction error compared with the known link method. The confidence bounds for the link function, given in Fig. 3(d), clearly exclude the linear link.

Data analysis and inference based on an assumed link function, although convenient, could lead to erroneous results. Our proposed methodology can suggest suitable candidates for the link function that can be explored by the data analyst and therefore is useful as a diagnostic tool.

**Acknowledgement**

The authors thank the editor, associate editor and two referees for their suggestions that led to several improvements of the article.

**Supplementary Material**

Supplementary material available at *Biometrika* online includes details of the proof of Theorem 1 and Theorem 2.

**Appendix**

To facilitate our arguments, we impose the following technical conditions.

*Condition A1.* The link function $g$ and the coefficient functions $\beta_k(\cdot)$ $(k = 1, \ldots, p)$ are three times continuously differentiable and $g$ is nonconstant on the support of $\theta^T X$.

*Condition A2.* The point $\theta_0 = (\beta_1(u_0), \ldots, \beta_p(u_0))^T$ is an interior point of a compact set $\Theta$.

*Condition A3.* The coefficient functions satisfy the identifiability constraint $\beta_1(u) > 0$ and $\sum_{k=1}^p \beta^2_k(u) = 1$ for $0 \leq u \leq 1$.

For each given $U = u_0$, with an unspecified link, model (1) is a single index model. This condition is the standard restriction (Ichimura, 1993; Lin & Kulasekera, 2007). It ensures that our objective function (3) has a well separated minimum (Van der Vaart, 1998) in the neighbourhood of the true parameter.

*Condition A4.* There exists a positive definite matrix $M_2(\theta_0)$ such that

$$E \left\{ \frac{\partial^2 M_n(\theta)}{\partial \theta \partial \theta^T} \right\}_{\theta = \theta_0} \to M_2(\theta_0), \quad n \to \infty,$$

where $M_2(\theta_0)$ is analogous to the information matrix in classical linear models.

This condition is similar to condition M7 of Chiou & Müller (1998) and condition 3 of Lemma 5.4 in Ichimura (1993).

*Condition A5.* As $n \to \infty$, $h_1 \sim n^{-\delta_1}$, $h_2 \sim n^{-\delta_2}$ with $0 < \delta_1 \leq 1/5 < \delta_2 < 1$.

*Condition A6.* The response variable is continuous with $E(\{Y\}^m) < \infty$, for some $m > 1 + (1 + 3\delta_1 + 2\delta_2)/(1 - 3\delta_1 - \delta_2)$. If we set the smoothing parameter $h_1 \sim n^{-1/2}$, which is the optimal order for nonparametric regression function estimators, we then require $m > 6$. We further assume the covariate $X = (x_1, \ldots, x_p)^T$ satisfies $\max_{1 \leq k \leq p} |x_k| \leq 1$.

*Condition A7.* The kernel functions $K_1(\cdot)$, $K_2(\cdot)$ are symmetric densities that are supported on $[-1, 1]$ and are continuously differentiable.
**Condition A8.** Let $A_n^k(X, \theta)$ be the $k$th partial derivative of $A_{n, u_0}(X, \theta)$ with respect to $\theta$ and let $A^k(X, \theta)$ be its probability limit. We assume $\sup_{(X, \theta) \in X \times \Theta} A_n^k(X, \theta) < \infty$ for $k = 0, 1, 2$ and $\inf_{t, \theta} f_{U|X, U}(t, u_0) > 0$.

**Condition A9.** Let
\[
\psi(u) = E[g'(\theta^*_0)g_1(X, \theta_0) \otimes X^t | U = u],
\]
\[
\rho(u, x) = E[(Y - g(\theta^*_0)X)^2 | U = u, X = x],
\]
\[
\Delta(u) = E[\rho(U, X)g_1(X, \theta_0)g_1(X, \theta_0)^T | U = u],
\]
where $\otimes$ denotes the Kronecker product. Here $\hat{g}^{(k)}(X, \theta)$ is the $k$th partial derivative of $\hat{g}_u(\theta^*_0)X, \theta)$ with respect to $\theta$ and $g_k(X, \theta)$ is its probability limit. We assume $\psi(u), \Delta(u), E(|Y|^3 | U = u)$ and the marginal density $f_U(\cdot)$ of $U$ are twice differentiable and $f_U(u_0) > 0$.

First we establish a uniform consistency result for our link function estimator that will be used in proving the asymptotic properties of the estimated varying coefficient functions. The proofs of the following lemma, along with the details of the proof of Theorems 1 and 2, are given in the online Supplementary Material.

**Lemma A1.** Under Conditions A1–A8, for $k = 0, 1, 2$,
\[
\sup_{(X, \theta) \in X \times \Theta} \left| \hat{g}^{(k)}(X, \theta) - g_k(X, \theta) \right| = o_p(1).
\]

**Proof of Theorem 2.** A Taylor series expansion of (3) yields
\[
0 = M_1^{(1)}(\hat{\theta}_0) + M_1^{(2)}(\hat{\theta}_0)(\hat{\theta} - \theta_0),
\]
where $\hat{\theta}$ between $\hat{\theta}$ and $\theta_0$ and $M_n^{(k)}(\theta^*)$ is the $k$th partial derivative of $M_n$ with respect to $\theta$ evaluated at $\theta = \theta^*$. Using $M_2(\theta_0)$ defined in Condition A4 and for a normalizing sequence $(nh_2)^{1/2}$, we can write (A3) as
\[
M_2(\theta_0)(nh_2)^{1/2}(\hat{\theta} - \theta_0) = -(nh_2)^{1/2}M_1^{(1)}(\theta_0) + [M_2(\theta_0) - M_1^{(2)}(\hat{\theta})](nh_2)^{1/2}(\hat{\theta} - \theta_0)
\]
\[
= \sum_{k=1}^{4} T_{1,k} + T_2.
\]

where
\[
T_{1,1} = d_n \sum_{i=1}^{n} |Y_i - g_0(X_i, \theta_0)|g_1(X_i, \theta_0)K_2 \left( \frac{U_i - u_0}{h_2} \right),
\]
\[
T_{1,2} = d_n \sum_{i=1}^{n} |Y_i - g_0(X_i, \theta_0)|\hat{g}^{(1)}(X_i, \theta_0) - g_1(X_i, \theta_0)K_2 \left( \frac{U_i - u_0}{h_2} \right),
\]
\[
T_{1,3} = d_n \sum_{i=1}^{n} (g_0(X_i, \theta_0) - \hat{g}_u(\theta^*_0X_i, \theta_0))g_1(X_i, \theta_0)K_2 \left( \frac{U_i - u_0}{h_2} \right),
\]
\[
T_{1,4} = d_n \sum_{i=1}^{n} (g_0(X_i, \theta_0) - \hat{g}_u(\theta^*_0X_i, \theta_0))\hat{g}^{(1)}(X_i, \theta_0) - g_1(X_i, \theta_0)K_2 \left( \frac{U_i - u_0}{h_2} \right),
\]
with $d_n = 2(nh_2)^{-1/2}$. Using Lemma A1, for suitably chosen bandwidth sequences $h_1$ and $h_2$ that satisfy Condition A5 and $nh_2 \to \infty$, we can easily show that $T_{1,k}$ ($k = 2, 3, 4$) converges in probability to zero as $n \to \infty$. It remains to show $T_{1,1}$ is asymptotically normal. Since $T_{1,1}$ is a sum of independent random vectors, asymptotic normality of $T_{1,1}$ follows from the Cramer–Wold device if we show that for any unit
vector $a$, $a'T_{1,1}$ converges to a univariate normal random variable. Hence, we will find the mean and the covariance matrix of $T_{1,1}$, and verify Lyapounov’s condition for the sequence $a'T_{1,1}$.

Let $\mu_j = \int s^j K_2(s) ds$, $\beta^{(1)}(u_0) = \{\beta_1^j(u_0), \ldots, \beta_p^j(u_0)\}^T$ and $\beta^{(2)}(u_0) = \{\beta_1''(u_0), \ldots, \beta_p''(u_0)\}^T$. Also let $\psi^{(1)}(u_0)$ be the $p \times p$ matrix of first derivatives of (A1) with respect to $u$ evaluated at $u_0$. Set

$$b(u_0) = f_U(u_0) \left\{ \psi^{(1)}(u_0) \beta^{(1)}(u_0) + \frac{1}{2} \psi(u_0) \beta^{(2)}(u_0) \right\} + f'_U(u_0) \{ \psi(u_0) \beta^{(1)}(u_0) \}.$$  \hspace{1cm} (A5)

Using (A1), (A2) and (A5) together with Conditions A8 and A9, standard calculations show that

$$E(T_{1,1}) = (nh_2)^{1/2} \{ 2\mu_2 h_2^2 b(u_0) + O(h_2^3) \}$$  \hspace{1cm} (A6)

and

$$\text{var}(T_{1,1}) = 4 f_U(u_0) v_0 \Delta(u_0) + o(1).$$  \hspace{1cm} (A7)

Now for any unit vector $a$, we will verify Lyapounov’s condition for the sequence $a'T_{1,1}$. To this end, write $a'T_{1,1} = \sum_{i=1}^n d_i W_i^{(n)} a^T g_1(X_i, \theta_0)$ where

$$W_i^{(n)} = \{ Y_i - g_0(X_i, \theta_0) \} K_2 \left( \frac{U_i - u_0}{h_2} \right).$$

Let $s_n^2 = \text{var}(a'T_{1,1})$. Since $g_0(\cdot, \theta_0) = g(\cdot)$, from (A7) we have $s_n^2 = 4 f_U(u_0) v_0 a^T \Delta(u_0) a + o(1)$ and hence Lyapounov’s condition holds for $a'T_{1,1}$ if $\sum_{i=1}^n E[|d_i W_i^{(n)} a^T g_1(X_i, \theta_0)|^{2+\delta}] \to 0$ for some $\delta > 0$. From Conditions A8 and A9, we have $E(|W_i^{(n)} a^T g_1(X, \theta_0)|^2) = O(h_2)$ which leads to

$$\sum_{i=1}^n E[|W_i^{(n)} a^T g_1(X_i, \theta_0)|^2] = O(n^{-1/2} h_2^{-1/2}).$$

Therefore, if $nh_2 \to \infty$ with $nh_2^2 \to 0$, we have $T_{1,1}$ converges in distribution to a multivariate normal random variable by the Cramer–Wold device.

It remains to show that $T_2$ converges in probability to zero. From Lemma A1, we can show that

$$\{ M_2^{(2)}(\hat{\theta}) - M_2(\theta_0) \} \to 0$$  \hspace{1cm} (A8)

in probability for suitably chosen bandwidth sequences satisfying Condition A5. Now we will show $(nh_2)^{1/2} (\hat{\theta} - \theta_0) = O_p(1)$. For a $p$-dimensional vector $x$, let $|x|_\infty = \max_{1 \leq k \leq p} |x_k|$. Since $M_2(\hat{\theta}) \leq M_2(\theta_0)$, a Taylor series expansion, Theorem 1 and (A8) gives

$$\hat{\theta} - \theta_0 \approx 1/2 (\hat{\theta} - \theta_0)^T M_2(\theta_0) (\hat{\theta} - \theta_0) + o_p(|\hat{\theta} - \theta_0|_\infty^2).$$  \hspace{1cm} (A9)

Following Ichimura (1993), multiplying both sides of (A9) by $(nh_2)^{1/2} |\phi_0 - \theta_0|_\infty^{-2}$ and letting $c_n(\hat{\theta}) = (nh_2)^{1/2} (\phi_0 - \theta_0) (1 + (nh_2)^{1/2} |\phi_0 - \theta_0|_\infty)^{-1}$ yields

$$c_n(\hat{\theta})^T (nh_2)^{1/2} M_2^{(1)}(\theta_0) + \frac{1}{2} c_n(\hat{\theta})^T M_2(\theta_0) c_n(\hat{\theta}) + \frac{o_p(1)}{1 + (nh_2)^{1/2} |\phi_0 - \theta_0|_\infty^2} \leq 0.$$  \hspace{1cm} (A10)

If $(nh_2)^{1/2} |\phi_0 - \theta_0|_\infty \to \infty$ in probability, then we have $c_n(\hat{\theta})$ converging in probability to a finite vector with at least one of the entries being equal to one. We have already shown that $(nh_2)^{1/2} M_2^{(1)}(\theta_0)$ converges in distribution to a finite random variable. Therefore, as $n \to \infty$ from (A10) we get

$$(1/2)c_n(\hat{\theta})^T M_2(\theta_0) c_n(\hat{\theta})\leq o_p(1).$$  \hspace{1cm} (A11)

Since $M_2(\theta_0)$ is positive definite (A11) leads to $c_n(\hat{\theta})$ converging to the zero vector in probability which is a contradiction. Hence, we must have $(nh_2)^{1/2} |\phi_0 - \theta_0|_\infty = O_p(1)$ which implies that $T_2 \to 0$ in probability as $n \to \infty$. Therefore, as $n \to \infty$ from (A4) we have $M_2(\theta_0)(nh_2)^{1/2} (\hat{\theta} - \theta_0)$ converging in distribution to a multivariate normal random variable. \hfill $\Box$
REFERENCES


*[Received February 2010. Revised April 2011]*