Nonparametric Threshold Regression: Estimation and Inference*

Daniel J. Henderson, Christopher F. Parmeter, and Liangjun Su

Abstract

The present work describes a simple approach to estimating the location of a threshold/change point in a nonparametric regression. This model has connections both to the time-series and regression discontinuity literatures. The estimator leverages a simple decomposition, giving it the form of a semiparametric smooth coefficient model. Optimal bandwidth selection and a suite of testing facilities are also presented. Several empirical examples are provided to illustrate the implementation of the methods discussed here.

Keywords: Change Point, Local Average Treatment Effect, Nonparametric Threshold Regression, Regression Discontinuity, Smoothed Bootstrap, Structural Change

JEL Codes:

1 Introduction

Regression discontinuity and structural change models have received considerable attention in the statistics and econometrics literature. There is well documented evidence of structural change in many economic time series, including GDP (McConnel and Perez-Quiros, 2000) and labor productivity (Hansen, 2001), along with change points in economic growth (Durlauf and Johnson, 1996; Hansen 2000) not to mention a myriad studies deploying regression discontinuity designs, where the change point is known to the analyst. A majority of the literature focuses on parametric models, though recently attention has shifted to detecting the presence of a structural change or the magnitude of a change point deploying nonparametric methods. While parametric methods possess the advantage of parsimony, the potential to avoid model misspecification through a nonparametric specification is alluring. However, many of the existing nonparametric methods for detecting structural breaks involve a diagnostic test or the use of one-sided kernels to estimate the unknown function on each side of the threshold. Here we describe a simple method to not only estimate the location of a structural change, but the unknown conditional mean on each side of the break.

Our method leverages a simple decomposition owing to the discrete nature of the structural change. This decomposition is identical to that appearing in Das (2005) albeit for a different econometric problem.

*Christopher F. Parmeter, Corresponding Author, Department of Economics, University of Miami, 305-284-4397, e-mail: cparmeter@bus.miami.edu.

However, both models, after the decomposition, take the form of a semiparametric smooth coefficient model,[1] which has been judiciously explored (see Li, Huang, Li and Fu, 2002; Lee and Ullah, 2001; Cai, Fan and Li, 2000; Cai, Fan and Yao, 2000; Fan and Huang, 2005; Cai, 2007; Cai, Li and Park, 2009; and Li and Racine, 2009, among others). The proposed estimator adds to a growing literature on threshold/change point estimation in nonparametric settings.

There currently exist several alternative approaches to estimating change points in nonparametric settings. Müller (1992) suggested estimating the location of a change point in an otherwise smooth regression surface by taking the maximal difference over all one-sided estimates of the unknown function on each side of the unknown change point. The estimates were constructed using one-sided kernels, similar to those deployed in boundary modification for the local constant estimator (Rice, 1984). This estimator was also used in Delgado and Hidalgo (2000) and Grégoire and Hamrouni (2002). Spokoiny (1998) uses a similar idea, constructing intervals over the data, estimating via local polynomial the unknown regression function, and then testing whether the residuals represent pure noise. The estimator is that which stems from the largest interval where this hypothesis cannot be rejected. An alternative approach to estimating the change point is to use a two-step approach. This is the route followed in Gijbels, Hall and Kniep (1999). The first step (the diagnostic step) involves looking for locations in the support of the data with estimated high derivatives for the local constant kernel regression estimator. The second step (the least squares step) fits a step function over a range of the data near the estimated discontinuities and uses least squares to determine the index of the data where the discontinuity is closest.

Gao, Gijbels and Van Bellegem (2008) test for structural breaks in a nonparametric location-scale model where both the conditional mean and variance functions may possess change points. Gao et al. (2008) generalize the approach of Hidalgo (1995) and Delgado and Hidalgo (2000) who require that the conditional mean and variance have the same location of the structural break. Seo and Linton (2006) generalize Hansen's (2000) change point regression model by allowing the threshold to be a linear index, as opposed to a single value.

Porter (2003) provides two estimators of the regression discontinuity (RD) treatment effect. The first estimator is based on Robinson's (1988) partially linear estimator (PLE). The second estimator is based on the local polynomial estimator (LPE) at the boundary which generalizes the local linear estimator of Hahn et al. (2001). Yu (2010) develops the partially polynomial estimator (PPE) which is a generalization of the PLE and connects the PLE and LPE of Porter (2003). Porter and Yu (2011) discuss estimation of the RD model when treatment assignment is unknown. This is an important distinction given that in RD models a concern is selection on treatment, where individuals react to the known cut-off to obtain treatment status, thus eliminating random assignment.

An important issue that arises when dealing with nonparametric estimation of a regression curve with jump discontinuities or kinks is the selection of the smoothing parameters. The performance of the

[1] Kristensen (2012) proposes a similar model to that found here, however, his model is setup as a semiparametric smooth coefficient model and the coefficients depend exclusively on time, whereas our nonparametric change point model is fully nonparametric in all variables, and takes the form of the smooth coefficient model through the discrete nature of the change point.

estimator depends crucially on the values of these parameters. However, few of the papers discussing change point estimation focus on selection of the bandwidth parameters. As Porter (2003) notes, the current crop of papers "…provide no practical guide to bandwidth choice …one could imagine using a leave-one-out cross-validation criterion evaluated at points outside a bandwidth neighborhood of the discontinuity," though no formal approach is given. Wu and Chu (1993) and Spokoiny (1998) are early contributions on the theoretical underpinnings of smoothing parameter selection. Gijbels and Goderniaux (2004) propose optimal bandwidth selection for the change point estimator of Gijbels et al. (1999), estimating the necessary bandwidths for the procedure using bootstrap bandwidth selection, while the number of discontinuities is determined via classic least-squares cross-validation. More recently Porter and Yu (2011) and Yu (2010) propose cross-validation algorithms for their estimators.

A new strand of analysis has focused on the deployment of wavelets to estimate structural breaks and change points in nonparametric regression models. Chen (2011) and Chen and Fan (2011) use local polynomial wavelets to estimate a local average treatment effect (LATE) in a switching regression with discontinuous incentive assignment. These estimators are constructed under the assumption of a known change point.

The objectives of this paper are twofold. We develop a competing nonparametric threshold model, using recently developed discrete smoothing methods. This method is simple to use and readily admits a simple cross-validation approach for automatic bandwidth selection. Further, we provide theoretical justification for our estimator, our bandwidth selection mechanism as well as the testing facilities. We also discuss how our estimator can be used in the RD treatment effect context.

The remainder of our paper is structured as follows. Section 2 presents our estimator, bandwidth selection algorithm and tests of several important hypotheses. Section 3 provides the theoretical underpinnings for our new estimator. Section 4 presents Monte Carlo evidence on the performance of our estimator. Several examples appear in Section 5. Section 6 concludes with several avenues for future research.

To proceed, we adopt the following notation. For a real matrix A, we denote its transpose as $A^{\intercal}$, its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA^{\intercal})]^{1/2}$), where $\mathrm{tr}(\cdot)$ is the trace operator and $\equiv$ means "is defined as". Let $1\{\cdot\}$ denote the usual indicator function that takes value if the condition insider $\{\cdot\}$ holds and zero otherwise. We use $\overset{D}{\to}$ and $\overset{P}{\to}$ to denote convergence in distribution and probability, respectively.

2 Threshold Regression

2.1 Parametric Threshold Regression

Consider the basic threshold regression model:

$$Y_t = \beta_1^\intercal X_t + \varepsilon_t, \quad q_t \leq \gamma, \tag{2.1}$$

$$Y_t = \beta_2^\intercal X_t + \varepsilon_t, \quad q_t > \gamma, \tag{2.2}$$

where X_t is a $d \times 1$ vector of regressors, ε_t is the error term, q_t is the threshold variable and is used to split the sample into two distinct regimes, and γ is referred to as the threshold parameter: the coefficient of X_t takes value β_1 when $q_t \leq \gamma$ and β_2 otherwise. In practice this model can be as simple as the variable q_t representing gender or race and γ is known to be 0 in the cross sectional setting, or it could be that the variable q_t is continuous and the value of γ is unknown, and needs to be estimated.

To think of an estimator for this model, we rewrite equations (2.1) and (2.2) into a single equation. In order to accomplish this we introduce the binary variable $D_t(\gamma) = 1\{q_t > \gamma\}$. This yields

$$Y_t = \beta_1^\intercal X_t + \delta^\intercal X_t D_t(\gamma) + \varepsilon_t, \tag{2.3}$$

where $\delta = \beta_2 - \beta_1$. This expression is actually a simplification since it is possible to have a model where a subset of variables have the same response effect across regimes or only belong to a single regime. However, writing the threshold regression in single equation form helps to see how generic least squares estimation applies to the estimation of the model. Writing (2.3) in matrix form we obtain

$$Y = X\beta_1 + \left(X \odot \bar{D}(\gamma)\right)\delta + \varepsilon, \tag{2.4}$$

where Y is the $n \times 1$ vector of regressands, ε is the $n \times 1$ vector of model errors, X is the $n \times d$ matrix of regressors, $\bar{D}(\gamma) = D(\gamma)\mathbf{1}_{1\times d}$, $D(\gamma)$ is the $n \times 1$ vector of indicators $D_t(\gamma)$ regarding regime assignment, $\mathbf{1}_{1\times d}$ a $1 \times d$ vector of ones, and $\odot$ denotes the Hadamard product. For a given level of γ, (2.4) can be solved using ordinary least squares for estimators of β_1 and δ. This stems from minimizing the sum of squared errors:

$$RSS(\beta_1, \delta|\gamma) = \left[Y - X\beta_1 - \left(X \odot \bar{D}(\gamma)\right)\delta\right]^\intercal \left[Y - X\beta_1 - \left(X \odot \bar{D}(\gamma)\right)\delta\right]. \tag{2.5}$$

In practice, γ is generally unknown. In this setting Hansen (2000) suggests obtaining estimates of $(\beta_2, \delta, \gamma)$ by concentration. That is, as shown in (2.5), conditional on γ the estimators for β_1 and δ are linear. Letting $\tilde{X}(\gamma) = [X\ X \odot \bar{D}(\gamma)]$, the concentrated OLS estimator of $(\beta^\intercal, \delta^\intercal)^\intercal$ is

$$(\hat{\beta}_1(\gamma)^\intercal, \hat{\delta}(\gamma)^\intercal)^\intercal = \left[\tilde{X}(\gamma)^\intercal \tilde{X}(\gamma)\right]^{-1} \tilde{X}(\gamma)^\intercal Y.$$

An estimator for γ can be found by defining $\hat{\gamma}$ as

$$\hat{\gamma} = \underset{\gamma\in\Gamma}{\arg\min} RSS(\hat{\beta}_1(\gamma), \hat{\delta}(\gamma)|\gamma), \tag{2.6}$$

where $\Gamma = [\underline{\gamma}, \bar{\gamma}]$. Hansen (2000) suggests approximating Γ with a grid and when n is large, one can use $N < n$ points to aid in computation. Large sample properties as well as inferential procedures are discussed in Hansen (2000) while the possibility to allow q to be endogenous is discussed in Caner and Hansen (2002). For the analysis that follows we will assume that q is exogenous.

2.2 A Nonparametric Threshold Regression

Now we consider the basic threshold regression model but leave the functional form unspecified:

$$Y_t = m_1(X_t) + \varepsilon_t, \quad q_t \leq \gamma, \tag{2.7}$$

$$Y_t = m_2(X_t) + \varepsilon_t, \quad q_t > \gamma, \tag{2.8}$$

where, $t = 1, ..., n$, q_t is again the threshold variable and is used to split the sample into two distinct regimes, $m_1(\cdot)$ and $m_2(\cdot)$ are two unknown smooth functions defined in $\mathbb{R}^d$. Now, we rewrite the nonparametric threshold regression estimator in single equation form, using the same binary variable introduced earlier, $D_t(\gamma)$. This yields

$$\begin{aligned} Y_t &= m_1(X_t) + [m_2(X_t) - m_1(X_t)] D_t(\gamma) + \varepsilon_t \\ &= \alpha_1(X_t) + \alpha_2(X_t) D_t(\gamma) + \varepsilon_t \end{aligned} \tag{2.9}$$

where $\alpha_1(X_t) = m_1(X_t)$ and $\alpha_2(X_t) = m_2(X_t) - m_1(X_t)$. This model is known as a semiparametric smooth coefficient model (SPSCM) and it has been extensively discussed in the econometric literature for the case where q_t is not an element of X_t or $D_t(\gamma)$ is replaced by another variable that is not a function of X_t. As above, for a fixed γ, the estimators of $\alpha_1(x)$ and $\alpha_2(x)$ can be obtained by minimizing the sum of squared residuals. This process can be iterated for a fixed grid Γ to obtain an estimator of the threshold parameter as well. Below, we will use subscript 0 to denote the true function or parameter value, e.g., γ_0 denotes the true value of γ and $\alpha_0(x) \equiv (\alpha_{1,0}(x), \alpha_{2,0}(x))^\intercal$ denotes the true function of $\alpha(x) \equiv (\alpha_1(x), \alpha_2(x))^\intercal$. Therefore the data generating process is given by

$$Y_t = \alpha_{1,0}(X_t) + \alpha_{2,0}(X_t) D_t(\gamma_0) + \varepsilon_t \tag{2.10}$$

where $\alpha_{1,0}(X_t) = m_{1,0}(X_t)$ and $\alpha_{2,0}(X_t) = m_{2,0}(X_t) - m_{1,0}(X_t)$ and $m_{1,0}$ and $m_{2,0}$ denote the true functions of m_1 and m_2.

Unfortunately, $(\alpha_{1,0}, \alpha_{2,0})$ is not identified if q_t is contained in X_t. Without loss of generality (wlog) we assume that $q_t = X_{1t}$, the first element of X_t. An alternative representation for the model in (2.9) is given by

$$Y_t = \alpha(X_t) + \beta(X_t) \cdot 1\{X_{1t} > \gamma\} + \varepsilon_t \tag{2.11}$$

where

$$\alpha(x) = \begin{cases} m_1(x) & \text{if } x_1 \leq \gamma \\ m_2(x) - \beta(x) & \text{if } x_1 \geq \gamma \end{cases}.$$

Note that $\alpha(x) = m_1(x) = m_2(x) - \beta(x)$ at $x_1 = \gamma$, ensuring the continuity of the function α at $x_1 = \gamma$. We can further require $\beta(x) = \beta$, a constant, and then obtain the following representation:

$$Y_t = \alpha(X_t) + \beta \cdot 1\{X_{1t} > \gamma\} + \varepsilon_t \tag{2.12}$$

where

$$\alpha(x) = \begin{cases} m_1(x) & \text{if } x_1 \leq \gamma \\ m_2(x) - \beta & \text{if } x_1 \geq \gamma \end{cases}.$$

In the last case, β can be interpreted as the jump size of the regression function at $x_1 = \gamma$. Despite the fact that $(\alpha(\cdot), \beta(\cdot))$ is not identified in (2.11), we can identify $((\beta,\gamma), \alpha(\cdot))$ under the condition that $\alpha(\cdot)$ is continuous everywhere on its support.

It is worth mentioning that the model in (2.9) or (2.12) is a nonparametric extension of the parametric regression discontinuity design (RDD) where γ is typically assumed to be known in the parametric framework. Here, we allow that the threshold parameter γ to be unknown.

In the following, we assume that $q_t = X_{1t}$ is a continuous random variable that admits a probability density function (PDF) f_1 and strictly increasing cumulative distribution function (CDF) F_1. We will propose estimates for (β, γ) and $m(x)$ and then establish their asymptotic distributions below. As one can imagine, like the parametric case the threshold parameter γ can be estimated at a rate faster than $\sqrt{n}$. Unlike the parametric case, the estimation of γ generally affects the asymptotic distribution of the estimator of β and may or may not asymptotic the asymptotic distribution of $\alpha(x)$, contingent upon whether $d = 1$ or $d > 1$.

When $q_t \notin X_t$, (2.9) becomes the standard functional functional model in the case where γ is known. When γ is unknown, similar but much simpler analysis than that in the current paper reveals the following results: 1) γ can be estimated at a rate fast than $\sqrt{n}$; 2) The estimation of γ does not have any first-order asymptotic effect on the asymptotic distribution of estimates of $m_1(x)$ and $m_2(x)$ as in the parametric case.

2.3 Semiparametric M-estimation of the threshold parameter

We we consider the semiparametric estimation of both the infinite dimensional parameter $\alpha(\cdot)$ and the finite parameter (β, λ) in (2.12). The associated DGP is

$$Y_t = \alpha_0(X_t) + \beta_0 \cdot 1\{X_{1t} > \gamma_0\} + \varepsilon_t \tag{2.13}$$

Like Hansen (2000), we estimate the γ by concentrating both $\alpha(\cdot)$ and β out. Our estimates can be obtained through a three-stage procedure.

1. In the first stage, for given (β, γ) and x we can estimate $\alpha(x)$ by Nadaraya-Watson (NW hereafter) method. The NW estimate of $\alpha(\cdot)$ is obtained as

$$\hat{\alpha}_b(x;\beta,\gamma) \equiv \arg\min_{\alpha} n^{-1} \sum_{t=1}^{n} [Y_t - \alpha - \beta 1\{q_t > \gamma\}]^2 K_b(X_t - x) \tag{2.14}$$

   where $K_b(X_t - x) = b^{-d} K(X_t - x)$ and $K(\cdot)$ is a kernel function. It is easy to verify that

$$\hat{\alpha}_b(x;\beta,\gamma) = n^{-1} \sum_{t=1}^{n} K_b(X_t - x)[Y_t - \beta 1\{q_t > \gamma\}] / \hat{f}_b(x), \tag{2.15}$$

   where $\hat{f}_b(x) = n^{-1} \sum_{t=1}^{n} K_b(X_t - x)$.

2. In the second stage, we choose β to minimize the following weighted least squares (WLS) objective function

$$n^{-1} \sum_{t=1}^{n} [Y_t - \hat{\alpha}_b(X_t;\beta,\lambda) - \beta 1\{q_t > \gamma\}]^2 \hat{f}_b^2(X_t). \tag{2.16}$$

The minimizer is given by

$$\hat{\beta}(\gamma) = \left(n^{-1} \sum_{t=1}^{n} \tilde{D}_t(\gamma)^2 \right)^{-1} n^{-1} \sum_{t=1}^{n} \tilde{D}_t(\gamma) \tilde{Y}_t, \tag{2.17}$$

where $\tilde{Y}_t = n^{-1} \sum_{s=1}^{n} K_b(X_s - X_t)(Y_s - Y_t)$ and $\tilde{D}_t(\gamma) = n^{-1} \sum_{s=1}^{n} K_b(X_s - X_t)[D_s(\gamma) - D_t(\gamma)]$. Let $\hat{\alpha}_b(x; \gamma) = \hat{\alpha}_b(x; \hat{\beta}(\lambda), \gamma)$.

3. In the third stage, we consider the semiparametric M-estimation of the threshold parameter γ. We estimate γ by $\hat{\gamma}$ that *approximately* solves the sample minimization problem:

$$\min_{\gamma \in \Gamma} \left| M_n(\gamma, \hat{h}_b) \right| \tag{2.18}$$

where $\hat{h}_b(\cdot; \gamma) = (\hat{\alpha}_b(\cdot; \gamma), \hat{\beta}(\gamma))$,

$$M_n(\gamma, \hat{h}_b) = \frac{1}{n} \sum_{t=1}^{n} \left[Y_t - \hat{\alpha}_b(X_t; \gamma) - \hat{\beta}(\gamma) D_t(\gamma) \right] w(X_t), \tag{2.19}$$

and $w(\cdot)$ is a nonnegative weight function with compact support $\mathcal{X}_0$ that lies in the interior of the support $\mathcal{X}$ of X_t. After one obtains $\hat{\gamma}$, one estimates β by $\hat{\beta}(\hat{\gamma})$ and $\alpha(x)$ by $\hat{\alpha}_b(x; \hat{\gamma}) = \hat{\alpha}_b(x; \hat{\beta}(\hat{\lambda}), \hat{\gamma})$.

Several remarks are in order.

First, we consider the NW estimation in this paper. Alternatively, one can consider other semiparametric estimation methods, e.g., local polynomial estimation and sieve estimation. The general results will be similar to what we have obtained in this paper.

Second, we consider the density-weighted least squares problem in second stage. The use of the estimated density as a weight helps to avoid the random denominator problem associated with NW estimation.

Third, we use the weight function $w(\cdot)$ in the third stage M-estimation and assume that it has compact support $\mathcal{X}_0$. This compact support assumption helps to trim the observations in the tail of $\mathcal{X}$ when $\alpha(\cdot)$ cannot be estimated accurately.

Fourth, note that M_n is not a smooth function of γ and we do not require $\hat{\gamma}$ to be the exact minimizer of the objective function in (2.15). As we shall see, our asymptotic theory requires that the *approximate* minimizer $\hat{\gamma}$ satisfies the condition

$$\left| M_n(\hat{\gamma}, \hat{h}_b) \right| = \inf_{\gamma \in \Gamma} \left| M_n(\gamma, \hat{h}_b) \right| + o_P\left((n/b)^{-1/2} \right), \tag{2.20}$$

where $(n/b)^{1/2}$ signifies the rate of convergence of $\hat{\gamma}$ to γ_0 under suitable conditions. Even though M_n is not a smooth function, it is a univariate function and one can easily obtain the approximate solution.

3 Asymptotic Properties

In this section we first study the asymptotic properties of the estimator $\hat{\gamma}$ and then consider the asymptotic properties of $\hat{\beta}(\hat{\gamma})$ and $\hat{\alpha}_b(x; \hat{\gamma})$.

3.1 Assumptions

For any $d \times 1$ vector $x = (x_1, ..., x_d)^\intercal$, we frequently write $x = (x_1, x_{-1}^\intercal)^\intercal$ where $x_{-1} = (x_2, ..., x_d)^\intercal$. For any vector $a = (a_1, ..., a_d)$ of d integers, define the differential operator $D^a = \partial D^{|a|} / \partial x_1^{a_1} ... \partial x_d^{a_d}$, where $|a| = \sum_{i=1}^d a_i$. For any $h : \mathcal{X}_0 \to R$ and $\lambda > 0$, let $\underline{\lambda}$ be the largest integer smaller than λ, and

$$\|h\|_{\infty,\lambda} = \max_{|a| \leq \underline{\lambda}} \sup_{x \in \mathcal{X}_0} |D^a h(x)| + \max_{|a| = \underline{\lambda}} \sup_{x \neq x'} \frac{|D^a h(x) - D^a h(x')|}{\|x - x'\|^{\lambda - \underline{\lambda}}}.$$

Let $C_c^\lambda(\mathcal{X}_0)$ be the set of all continuous functions $h : \mathcal{X}_0 \to R$ with $\|h\|_{\infty,\lambda} \leq c$ for some $c < \infty$.

We make the following assumptions.

Assumption A1. (i) The process $\{(X_t, \varepsilon_t)\}$ is a strictly stationary and β-mixing with mixing coefficients β_τ satisfying $\beta_\tau \leq c_\beta \rho^\tau$ for some $c_\beta > 0$ and $\rho \in (0, 1)$.

(ii) $E(\varepsilon_t | X_t, X_{t-1}, ..., \varepsilon_{t-1}, ...) = 0$ almost surely. $E|\varepsilon_t|^{2+\epsilon} < \infty$ for some $\epsilon > 0$.

(iii) The probability density function (PDF) $f(\cdot)$ of X_t is continuously differentiable, bounded, and bounded away from 0 on the compact subset $\mathcal{X}_0$ of its support $\mathcal{X}$. For all $l \geq 1$, the joint PDF $f_l(\cdot, \cdot)$ of X_t and X_{t+l} is uniformly bounded.

(iv) The conditional distribution function (CDF) $F_1(\cdot)$ of $q_t = X_{t1}$ admits a PDF $f_1(\cdot)$ that is uniformly bounded on its support.

Assumption A2. (i) Let $\upsilon \geq 2$ be an even integer. $f(\cdot)$ is υ-th order continuously differentiable on the compact set $\mathcal{X}_0$.

(ii) There exists $\lambda > d$ such that $\alpha_0(\cdot) \in C_c^\lambda(\mathcal{X}_0)$ for some $c > 0$. The $(\upsilon + \underline{\lambda})$th order partial derivatives of $\alpha_0(\cdot)$ exist and are continuous on $\mathcal{X}_0$.

(iii) The nonnegative weight function $w(\cdot)$ is second order continuously differentiable on is compact support $\mathcal{X}_0$. $\mathcal{X}_0$ is a product space and can be written as $\mathcal{X}_{0,1} \times \mathcal{X}_{0,-1}$, where $\mathcal{X}_{0,1}$ is a compact set on the real line that includes γ_0 as its interior point.

Assumption A3. (i) The kernel function $K(\cdot)$ is a product kernel of $k(\cdot)$ that is a symmetric υ-th order kernel with compact support $[-1, 1]$.

(ii) $k(\cdot)$ is $\underline{\lambda}$th order continuously differentiable with the $\underline{\lambda}$th order derivative $k^{(\underline{\lambda})}(\cdot)$ satisfying the Lipschitz condition $\left|k^{(\underline{\lambda})}(u) - k^{(\underline{\lambda})}(v)\right| \leq c_k |u - v|$ for all $u, v \in [-1, 1]$.

(iii) Let $\bar{k}(v) = \int_{-1}^v k(s) \, ds$ and $c_{\bar{k}} = 1 - 2\int_0^1 \bar{k}(v) \, dv + \int_{-1}^1 \bar{k}^2(v) \, dv$. Assume that $c_{\bar{k}} \neq |\dot{e}_w(\gamma_0)| / (4e_w(\gamma_0))$, where $e_w(x_1) = \int w(x) f(x) \, dx_{-1}$, and $\dot{e}_w(x_1) = \partial e_w(x_1) / \partial x_1$.

Assumption A4. (i) Assume that $b \propto n^{-\eta}$ for some η such that

$$\max\left(\frac{1}{2\upsilon + 1}, \frac{1}{d + 2\upsilon - 1}\right) < \eta < \min\left(\frac{1}{2d - 1}, \frac{1}{d + 2\underline{\lambda}}\right)$$

(ii) There exists $\lambda_0 \in (d, \min(2d, \lambda))$ such that $\frac{d\eta}{2d - \lambda_0} < \kappa < \min\left(\eta\upsilon, \frac{1}{2}(1 - d\eta)\right)$.

Assumption A1 imposes standard conditions on the stochastic process. We assume β-mixing instead of a weaker condition, α-mixing, because we will resort to some stochastic equicontinuity result established for β-mixing processes established in Doukhan et al. (1995). The geometric decay of β_t will facilitate the use of a Bernstein-type inequality for strong mixing processes. A2(i)-(ii) impose standard conditions on the unknown smooth functions, f and α_0, to ensure the uniform consistency of NW estimates on a compact set; see, e.g., Masry (1996) and Hansen (2008). The even integer υ denotes the order of the kernel function we use. A2(i)-(ii) are needed to ensure that the asymptotic bias of the kernel estimator of f and that of the kernel estimator of the $\underline{\lambda}$th derivative of $\alpha_0(\cdot)$ are both $O_P(b^\upsilon)$. A2(iii) assumes conditions on the weight function $w(\cdot)$.

Assumption A3 imposes conditions on the kernel function. We allow the use of higher order kernel to eliminate the asymptotic bias of the NW estimates. We assume compact support for the univariate kernel function $k(\cdot)$, which greatly facilitates the asymptotic analysis of our estimators. A3(ii) ensures the $\underline{\lambda}$th derivative of the estimate of $\alpha_0(x)$ to be well behaved. A3(iii) is needed to identify the threshold parameter γ. Assumption A4 mainly imposes conditions on the bandwidth sequence. A4(i) implies that

$$\max(nb^{2\upsilon+1}, nb^{d+2\upsilon-1}) \to 0 \text{ and } \min(nb^{2d-1}/(\log n)^2, nb^{d+2\underline{\lambda}}) \to \infty$$

and we must apply an undersmoothing bandwidth b in order to eliminate the effect of the asymptotic bias in the first stage nonparametric estimation on the second stage parameter estimation. It also implies that we have to resort a kernel function whose order is higher than $\underline{\lambda}$. For example, if $\underline{\lambda}= 2$, we need to choose $\upsilon \geq 4$ so that a higher order kernel has to be used. A4(ii) is used to establish some stochastic equicontinuity result which is stronger than what is typically needed in order to establish the usual $\sqrt{n}$-rate convergence for some parameter estimate. The reason is that our threshold parameter estimate $\hat{\gamma}$ has a rate of convergence faster than the usual $\sqrt{n}$-rate. Let $\kappa \in (\frac{d\eta}{2d-\lambda_0}, \min\left(\eta\upsilon, \frac{1}{2}(1-d\eta)\right))$. The condition that $\kappa < \min\left(\eta\upsilon, \frac{1}{2}(1-d\eta)\right)$ ensures that $b^\upsilon + n^{-1/2}b^{-d/2}(\log n)^{1/2} = o(n^{-\kappa})$ and that $\frac{d\eta}{2d-\lambda_0} < \kappa$ ensures that we can apply some Bernstein inequality to prove some stochastic equicontinuity results.

To appreciate Assumption A4 more, we focus on the case $d = 1$ and discuss two subcases.

1. $d = 1$ and $\lambda \in (1, 2]$. In this case, $\underline{\lambda}= 1$ and it suffices to consider second order kernel ($\upsilon = 2$). Then for any $\eta \in \left(\frac{1}{4}, \frac{1}{3}\right)$, all the conditions in A4 will be satisfied by restricting $\lambda_0 = \min(2 - \frac{\eta}{1-\eta} - \epsilon, \lambda)$ for any $\epsilon > 0$ such that $2 - \frac{\eta}{1-\eta} - \epsilon > 1$, which is possible because $\frac{\eta}{1-\eta} \in \left(\frac{1}{3}, \frac{1}{2}\right)$ when $\eta \in \left(\frac{1}{4}, \frac{1}{3}\right)$. That is, in this case, we can use a second order with an undersmoothing bandwidth $b \propto n^{-\eta}$ with $\eta \in \left(\frac{1}{4}, \frac{1}{3}\right)$.

2. $d = 1$ and $\lambda \in (2, 3]$. In this case, one can continue to apply the second order kernel with previously defined rate of bandwidth. Alternatively, we can apply a fourth order kernel ($\upsilon = 4$). Then for any $\eta \in \left(\frac{1}{8}, \frac{1}{5}\right)$ all the conditions in A4 will be satisfied by restricting $\lambda_0 = \min(2 - \frac{\eta}{1-\eta} - \epsilon, \lambda)$ for any $\epsilon > 0$ such that $2 - \frac{\eta}{1-\eta} - \epsilon > 1$, which is possible because $\frac{\eta}{1-\eta} \in \left(\frac{1}{7}, \frac{1}{4}\right)$ when $\eta \in \left(\frac{1}{8}, \frac{1}{5}\right)$. That is, in this case, we can use a fourth order with an undersmoothing bandwidth $b \propto n^{-\eta}$ with $\eta \in \left(\frac{1}{8}, \frac{1}{5}\right)$.

3.2 Consistency and asymptotic normality of $\hat{\gamma}$

To study the asymptotic properties of $\hat{\gamma}$, we first define the population analogue of $\hat{h}_b(x,\gamma) = (\hat{\alpha}_b(x,\gamma), \hat{\beta}_p(\gamma))^\intercal$. Let

$$\begin{aligned}\bar{\alpha}_b(x;\beta,\gamma) &\equiv \arg\min_{\alpha} E\left\{[Y_t - \alpha - \beta D_t(\gamma)]^2 K_{t,x}\right\},\\ \bar{\beta}_b(\gamma) &\equiv \arg\min_{\alpha} E\left[Y_t - \bar{\alpha}_b(X_t;\beta,\gamma) - \beta 1\{q_t > \gamma\}\right]^2 f^2(X_t),\\ \bar{\alpha}_b(x,\gamma) &\equiv \arg\min_{\alpha} E\left\{\left[Y_t - \alpha - \bar{\beta}_b(\gamma) D_t(\gamma)\right]^2 K_{t,x}\right\},\end{aligned}$$

where $K_{t,x} = K_b(X_t - x)$. Define

$$d_b(x,\gamma) \equiv \frac{1}{b} E\left\{K_{t,x}\left[D_t(\gamma) - 1\{x_1 > \gamma\}\right]\right\}, \text{ and} \tag{3.1}$$

$$c_{d_b}(\gamma) \equiv \left(E\left[d_b^2(X_t,\gamma)\right]\right)^{-1} E\{d_b(X_t,\gamma)\left[d_b(X_t,\gamma_0) - d_b(X_t,\gamma)\right]\}. \tag{3.2}$$

It is easy to verify that $\bar{\alpha}_b(x;\beta,\gamma) = \alpha_{0,b}(x;\beta,\gamma) + O(b^v)$, $\bar{\beta}_b(\gamma) = \beta_{0,b}(\gamma) + O(b^v)$, and $\bar{\alpha}_b(x,\gamma) = \alpha_{0,b}(x,\gamma) + O(b^v)$, where

$$\alpha_{0,b}(x;\beta,\gamma) = \alpha_0(x) + f(x)^{-1} E\{K_{t,x}\left[\beta_0 D_t(\gamma_0) - \beta D_t(\gamma)\right],\tag{3.3}$$

$$\beta_{0,b}(\gamma) = \beta_0 + \beta_0 c_{d_b}(\gamma), \text{ and}\tag{3.4}$$

$$\alpha_{0,b}(x,\gamma) = \beta_0 f(x)^{-1} E\{K_{t,x}\left[D_t(\gamma_0) - D_t(\gamma)\right]\} - \beta_0 c_{d_b}(\gamma) f(x)^{-1} E\left[K_{t,x} D_t(\gamma)\right].\tag{3.5}$$

Apparently, $\beta_{0,b}(\gamma_0) = \beta_0$ and $\alpha_{0,b}(X_t,\gamma_0) = \alpha_0(X_t)$. Noting that $d_b(x,\gamma) = \frac{1}{b}\int K(u)\,[1\left\{u_1 > -\frac{x_1-\gamma}{b}\right\}$ $-1\left\{0 > -\frac{x_1-\gamma}{b}\right\}]f(x+bu)\,du$, we can write $d_b(x,\gamma) = \bar{d}_b(x_1-\gamma;x)$, where $\bar{d}_b(t;x) = \frac{1}{b}\bar{d}\left(\frac{t}{b};x\right)$ and $\bar{d}(\cdot;x) = \int K(u)\left[1\{\cdot > -u_1\} - 1\{\cdot > 0\}\right] f(x+bu)\,du$ behaves like a univariate kernel function varying over x. One can verify that $E\left|d_b(X_t,\gamma)\right| = O(1)$, and $E[\left|d_b(X_t,\gamma)\right|^2] = O(1/b)$.

Let $h_b(\cdot,\gamma) \equiv (\alpha_b(\cdot,\gamma), \beta_b(\gamma))$ and $h_{0,b}(\cdot,\gamma) = (\alpha_{0,b}(\cdot,\gamma), \beta_{0,b}(\gamma))$. For notational convenience, we usually suppress the arguments of the function h_b and write $(\gamma, h_b) \equiv (\gamma, h_b(\cdot,\gamma))$, $(\gamma, h_{0,b}) \equiv (\gamma, h_{0,b}(\cdot,\gamma))$, and $(\gamma_0, h_{0,b}) \equiv (\gamma_0, h_{0,b}(\cdot,\gamma_0)) = (\gamma_0, (\alpha_0(\cdot), \beta_0))$. Define the pseudo-norm $\|\cdot\|_{\mathcal{H}}$ for h_b to lie in an infinite dimensional parameter set $\mathcal{H}$ to be defined in the appendix:

$$\|h_b\|_{\mathcal{H}} = \sup_{\gamma\in\Gamma}\sup_{x\in\mathcal{X}_0}\left|\alpha_b(x,\gamma)\right| + \sup_{\gamma\in\Gamma}\left|\beta_b(\gamma)\right|.$$

The following theorem establishes the consistency of $\hat{\gamma}$.

Theorem 3.1 *Suppose that Assumptions A1-A3 and A4(i) hold. Then* $\hat{\gamma} - \gamma_0 = o_P(1)$.

The proof of the above theorem is quite tedious as one cannot directly apply some existing results, e.g., Chen Linton, and Van Keilegom (2003, CLV hereafter), in the literature. To appreciate why, condition (1.2) in CLV requires that for any fixed $\delta > 0$, there exists $\epsilon(\delta)$ such that

$$\inf_{|\gamma-\gamma_0|>\delta}\left|M(\gamma, h_{0,b})\right| \geq \epsilon(\delta) > 0,$$

where $M(\gamma, h_b) \equiv E\{[Y_t - \alpha_b(X_t,\gamma) - \beta_b(\gamma)D_t(\gamma)]\,w(X_t)\}$. The above condition serves as a *strong* identification condition in the framework of CLV. It implies that

$$\Pr\left(|\hat{\gamma} - \gamma_0| > \delta\right) \leq \Pr\left(\left|M(\hat{\gamma}, h_{0,b})\right| \geq \epsilon(\delta)\right)$$

and thus one can prove the consistency of $\hat{\gamma}$ by showing that $|M(\hat{\gamma}, h_{0,b})| = o_P(1)$. Unfortunately, such a strong identification condition does not hold in our framework. In fact, for any fixed $\delta > 0$, there is no way to ensure that $|M(\gamma, h_{0,b})|$ is bounded away from zero uniformly in $\gamma \in \bar{\Gamma}_\delta \equiv \{\gamma' \in \Gamma : |\gamma' - \gamma_0| > \delta\}$. In other words, $M(\gamma, h_{0,b})$, as a function of γ, is quite flat in the neighborhood of γ_0, giving rise to the issue of *weak* identification. In the proof of the above theorem, we show that $\frac{1}{b}M(\gamma, h_{0,b})$ is bounded away from zero in the neighborhood of γ_0. The division of $M(\gamma, h_{0,b})$ by b helps to achieve the identification but it also causes some additional problems to be solved with care. For example, the standard consistency result in CLV requires that

$$\left\|\hat{h}_b - h_{0,b}\right\|_{\mathcal{H}} = o_P(1)$$

and

$$\sup_{\gamma \in \Gamma, \|h_b - h_{0,b}\|_{\mathcal{H}} \le \delta_n} |M_n(\gamma, h_b) - M(\gamma, h_b)| = o_P(1)$$

where $\delta_n = o(1)$ is an arbitrary positive sequence. We need to strengthen $o_P(1)$ to $o_P(b)$ in order to establish the claimed result in Theorem 3.1.

To state the next result, we introduce more notations. Define

$$\begin{aligned} c_{0b}(x, \gamma) &\equiv E[K_b(X_t - x) D_t(\gamma)], \\ c_{1b}(x, \gamma) &\equiv E[d_b(X_t, \gamma) K_b(X_t - x)], \\ c_{2b}(x, \gamma) &\equiv d_b(x, \gamma) E[K_b(X_t - x)], \\ \bar{c}_{0b}(\gamma) &\equiv \frac{1}{b}\int [c_{0b}(x, \gamma) - 1\{x_1 > \gamma\} f(x)] w(x) dx. \end{aligned} \tag{3.6}$$

It is easy to verify that $E|c_{sb}(X_t, \gamma)| = O(1)$ for $s = 0, 1, 2$, $E[|c_{sb}(X_t, \gamma)|^2] = O(1/b)$ for $s = 1, 2$, and $\bar{c}_{0b}(\gamma) = O(1)$. Let

$$V_{\gamma,b} \equiv \bar{c}_{0b}^2(\gamma_0) S_b^{-2}(\gamma_0) E\left\{\sigma^2(X_t) b [c_{1b}(X_t, \gamma_0) - c_{2b}(X_t, \gamma_0)]^2\right\} \tag{3.7}$$

where $\sigma^2(x) \equiv E(\varepsilon_t^2 | X_t = x)$, and $S_b(\gamma) = b \cdot E[d_b^2(X_t, \gamma)]$.

Let $\hat{\gamma}_+$ and $\hat{\gamma}_-$ denote $\hat{\gamma}$ when $\hat{\gamma} > \gamma_0$ and $\hat{\gamma} < \gamma_0$, respectively. Let $\Upsilon_{1b,-}(\gamma, h_{0,b})$ and $\Upsilon_{1b,+}(\gamma, h_{0,b})$ denote the ordinary left and right derivatives of $M(\gamma, h_{0,b})$ with respect to γ, respectively. We verify in Appendix B.4 that $\Upsilon_{1b,-}(\gamma, h_{0,b})$ and $\Upsilon_{1b,+}(\gamma, h_{0,b})$ exist for all γ in the neighborhood of γ_0 and $\Upsilon_{1b,-}(\gamma_0, h_{0,b})$ and $\Upsilon_{1b,+}(\gamma_0, h_{0,b})$ are both continuous at $\gamma = \gamma_0$ and bounded away from zero and infinity as $n \to \infty$. In particular, (B.20) gives the formula for $\Upsilon_{1b,\pm} \equiv \Upsilon_{1b,\pm}(\gamma_0, h_{0,b})$:

$$\Upsilon_{1b,\pm} = -\beta_0 e_w(\gamma_0) \pm \beta_0 \frac{1}{4c_{\bar{k}}} \dot{e}_w(\gamma_0) + O(b), \tag{3.8}$$

where $e_w(x_1)$, $\dot{e}_w(x_1)$, and $c_{\bar{k}}$ are defined in Assumption A3(iii).

The following theorem reports the asymptotic normality of $\hat{\gamma}$.

Theorem 3.2 *Suppose that Assumptions A1-A4 hold. Then*

$$\sqrt{n/b}\,(\hat{\gamma}_\pm - \gamma_0) \xrightarrow{D} N(0, \Omega_{\gamma,\pm})$$

where $\Omega_{\gamma,\pm} \equiv \lim_{n\to\infty} \Upsilon_{1b,\pm}^{-1} V_{\gamma,b} \Upsilon_{1b.\pm}^{-1}$, $\hat{\gamma}_\pm$ *denotes either* $\hat{\gamma}_+$ *or* $\hat{\gamma}_-$, *and similarly* $\Upsilon_{1b,\pm}$.

We make several remarks on the above theorem.

First, as in the case of parametric threshold regression, the threshold parameter in our nonparametric model can be estimated at a rate faster than the usual $\sqrt{n}$-rate. Chan (1993) finds that with the jump size parameter (β in our model) fixed, $n(\hat{\gamma}-\gamma_0)$ converges to an asymptotic distribution that is dependent upon some nuisance parameters and thus not particularly useful for statistical inference. For this reason, Hansen (2000) assumes that the jump size parameter is proportional to n^{-a} with $0<a<\frac{1}{2}$ and finds that $n^{1-2a}(\hat{\gamma}-\gamma_0)$ converges to asymptotic distribution that is associated with a two-sided Brownian motion. Note that the convergence rate of Hansen's threshold parameter estimate is faster than $\sqrt{n}$ provided $a<\frac{1}{4}$ so that the jump size does not shrink to zero too fast. We achieve the asymptotic normal distribution and faster convergence rate without assuming the jump size to shrink to zero.

Second, our result is similar to that in Seo and Linton (2007). The latter authors study the smoothed least squares estimation of a *parametric* threshold regression model where the indicator function ($1\{q_t>\gamma\}$ in our case) is replaced by a CDF-type smooth function with a bandwidth parameter σ to control the speed at which the CDF-type smooth function approximates the indicator function. They demonstrate that the threshold parameter in their model can be estimated at $\sqrt{n/\sigma}$-rate. Despite the similarity in terms of convergence rate for the threshold parameters, the asymptotic tools used in our paper is quite different from those used by Seo and Linton. The objective function in Seo and Linton (2007) is a smooth function so that they can apply the usual Taylor expansions whereas the objective function in our case is not smooth and we have to rely on the empirical process theory.

Third, the proof of Theorem 3.2 is quite tedious too as one cannot apply any existing results in the literature directly. For example, one cannot apply either the asymptotic normality result in CLV or that in Chen (2007) as they require that the ordinary derivative of $M(\gamma,h_{0,b})$ with respect to γ exists in the neighborhood of γ_0. In our model, we can only demonstrate that both the left and right derivatives of $M(\gamma,h_{0,b})$ exists at $\gamma=\gamma_0$. It turns that this condition, in conjunction with some other regularity conditions, is sufficient for the establishment of the asymptotic distribution of $\hat{\gamma}$. In addition, the stochastic equicontinuity (s.e.) condition (e.g., condition (2.5) in CLV and condition (4.1.5) in Chen (2007)) is not sufficient for our purpose either. We require a stronger s.e. condition than theirs and verify such a condition by relying upon some standard arguments (e.g., chaining argument) used in the empirical process theory.

Fourth, in principle, one can rely on the asymptotic results in Theorem 3.2 to make statistical inference about γ_0. To do so, one needs to estimate $\Upsilon_{1b,\pm}$ and $V_{\gamma,b}$ consistently. Let $m_w(x_1)=E[w(X_t)|X_{1t}=x_1]$. Noting that $e_w(x_1)=m_w(x_1)f(x_1)$, we propose to estimate $e_w(\gamma_0)$ by

$$\hat{e}_w=\frac{1}{nb}\sum_{t=1}^{n}k\left(\frac{X_{1t}-\hat{\gamma}}{b}\right)w(X_t)\ \text{ if } d\geq 2 \text{ and } w(\hat{\gamma})\hat{f}_b(\hat{\gamma})\ \text{ if } d=1,$$

and $\dot{e}_w(\gamma_0)$ by

$$\widehat{\dot{e}}_w=\frac{1}{nb^2}\sum_{t=1}^{n}k^{(1)}\left(\frac{\hat{\gamma}-X_{1t}}{b}\right)w(X_t)\ \text{ if } d\geq 2 \text{ and } \dot{w}(\hat{\gamma})\hat{f}_b(\hat{\gamma})+w(\hat{\gamma})\hat{f}_b^{(1)}(\hat{\gamma})\ \text{ if } d=1,$$

where $k^{(1)}(v)=dk(v)/dv$, $\dot{w}(\gamma)=\partial w(\gamma)/\partial\gamma$ when $d=1$, and $\hat{f}_b^{(1)}(\hat{\gamma})=\frac{1}{nb^2}\sum_{t=1}^{n}k^{(1)}\left(\frac{\hat{\gamma}-X_{1t}}{b}\right)$. Then by (3.8) we can estimate them respectively by

$$\hat{\Upsilon}_{1b,\pm} = -\hat{\beta}_b(\hat{\gamma})\,\hat{e}_w \pm \hat{\beta}_b(\hat{\gamma})\,\frac{1}{4c_{\bar{k}}}\hat{\dot{e}}_w.$$

Noting that by (B.18) and (B.19),

$$V_{\gamma,b} \equiv \left(\frac{\dot{e}_w(\gamma_0)}{2e(\gamma_0)\,c_{\bar{k}}}\right)^2 E\left\{\sigma^2(X_t)\,b\left[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)\right]^2\right\} + O(b)\,, \tag{3.9}$$

we propose to estimate $V_{\gamma,b}$ by

$$\hat{V}_{\gamma,b} = \left(\frac{\hat{\dot{e}}_w}{2\hat{e}c_{\bar{k}}}\right)^2 \frac{1}{n}\sum_{t=1}^{n}\hat{\varepsilon}_t^2 b\left[\hat{c}_{1b}(X_t,\hat{\gamma}) - \hat{c}_{2b}(X_t,\hat{\gamma})\right]^2$$

where $\hat{c}_{jb}(x,\hat{\gamma})$, $j = 1,2$, are sample analogue estimates of $c_{jb}(x,\hat{\gamma})$, and $\hat{\varepsilon}_t = Y_t - \hat{\alpha}_b(X_t,\hat{\gamma}) - \hat{\beta}_b(\hat{\gamma})\,D_t(\hat{\gamma})$. Then a consistent estimate of $\Omega_{\gamma,\pm}$ is given by $\hat{\Omega}_{\gamma,\pm} = \hat{V}_{\gamma,b}/(\hat{\Upsilon}_{1b,\pm})^2$. To test the null hypothesis $H_0 : \gamma = \gamma_0$, say, we can construct the t-statistic as usual

$$t_{n,\pm} = \sqrt{n/b}\left(\hat{\gamma}_\pm - \gamma_0\right)/\sqrt{\hat{\Omega}_{\gamma,\pm}}$$

where $t_{n,+}$ is used if $\hat{\gamma} > \gamma_0$ (in which case we write $\hat{\gamma}$ as $\hat{\gamma}_+$) and $t_{n,-}$ is used if $\hat{\gamma} < \gamma_0$ (in which case we write $\hat{\gamma}$ as $\hat{\gamma}_-$). Difficulty arises when one tries to construct the confidence interval for γ_0 as it is difficult to determine whether one should use $\hat{\Omega}_{\gamma,+}$ or $\hat{\Omega}_{\gamma,-}$. We propose to adopt the IID bootstrap to conduct the inference based on confidence intervals. Following the arguments used in Seo and Linton (2007), one can justify the asymptotic validity of this bootstrap method.)

3.3 Asymptotic distributions of $\hat{\beta}_b(\hat{\gamma})$ and $\hat{\alpha}_b(x;\hat{\gamma})$

Let $\theta_{\psi_1}(\gamma) = b\cdot E\{d_b(X_t,\gamma)\left[d_b(X_t,\gamma_0) - d_b(X_t,\gamma)\right]\}$. Let $\dot{\theta}_{\psi_1,+}(\gamma_0)$ and $\dot{\theta}_{\psi_1,-}(\gamma_0)$ denote the right and left derivatives of $\theta_{\psi_1}(\gamma)$ evaluated at $\gamma = \gamma_0$, respectively. We show in Appendix B that $b\cdot\dot{\theta}_{\psi_1,+}(\gamma_0) = -\frac{1}{2}e(\gamma_0) + O(b)$ and $b\cdot\dot{\theta}_{\psi_1,-}(\gamma_0) = \frac{1}{2}e(\gamma_0) + O(b)$, where $e(x_1) = \int f(x_1,x_{-1})^3\,dx_{-1}$. Define

$$\Omega_{n\beta,\pm} = S_b^{-2}(\gamma_0)\left[1 - \beta_0 b\dot{\theta}_{\psi_1,\pm}(\gamma_0)\,\Upsilon_{1b,\pm}^{-1}\bar{c}_{0b}(\gamma_0)\right]^2 E\{b\left[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)\right]^2\sigma^2(X_t)\}, \tag{3.10}$$

and

$$\Delta_{n\alpha,\pm}(x;d) = b^{(d-1)}c_{\alpha,b,\pm}^2(x)\,E\{b\left[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)\right]^2\sigma^2(X_t)\}, \tag{3.11}$$

where

$$c_{\alpha,b,\pm}(x) = f(x)^{-1}S_b^{-1}(\gamma_0)\{\beta_0 b\dot{c}_{0b}(x,\gamma_0)\,\bar{c}_{0b}(\gamma_0) - c_{0b}(x,\gamma_0)\left[1 - \beta_0 b\dot{\theta}_{\psi_1,\pm}(\gamma_0)\,\Upsilon_{1b,\pm}^{-1}\bar{c}_{0b}(\gamma_0)\right]\}, \tag{3.12}$$

and $\dot{c}_{0b}(x,\gamma) = \partial c_{0b}(x,\gamma)/d\gamma$.

The following theorem reports the asymptotic distributions of $\hat{\beta}_b(\hat{\gamma})$ and $\hat{\alpha}_b(x;\hat{\gamma})$.

Theorem 3.3 *Suppose that Assumptions A1-A4 hold. Suppose that $\sigma^2(\cdot)$ is continuous at x. Then*

(i) $\sqrt{nb}\left(\hat{\beta}_b(\hat{\gamma}_\pm) - \beta_0\right) \xrightarrow{D} N(0,\Omega_{\beta,\pm})$,

(ii) $\sqrt{nb^d}\left[\hat{\alpha}_b(x,\hat{\gamma}_\pm) - \alpha_0(x)\right] \xrightarrow{D} N\left(0, f(x)^{-1}\sigma^2(x)\int K(u)^2\,du + \Delta_{\alpha,\pm}(x;d)\right)$,

where $\Omega_{\beta,\pm} = \lim_{n\to\infty}\Omega_{n\beta,\pm}$, *and* $\Delta_{\alpha,\pm}(x;d) = \lim_{n\to\infty}\Delta_{n\alpha,\pm}(x;d)$.

Several remarks are in order.

First, although $\hat{\gamma}$ is super-consistent, $\hat{\beta}$ converges to β_0 at the nonparametric $\sqrt{nb}$-rate. This is due to the presence of $q_t = X_{1t}$ inside the indicator function $1\{q_t > \gamma\}$ but not the weak identification of γ_0. To appreciate this point, letting $\xi_n \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^n b^{1/2} [c_{1b}(X_t, \gamma_0) - c_{2b}(X_t, \gamma_0)] \varepsilon_t$, we show in the proof of Theorem 3.3(i) that

$$\sqrt{nb}\left(\hat{\beta}_b(\hat{\gamma}_{\pm}) - \beta_0\right) = S_b^{-1}(\gamma_0)\xi_n - \beta_0 b\dot{\theta}_{\psi_1,\pm}(\gamma_0)(b\Gamma_{1b,\pm})^{-1}\bar{c}_{0b}(\gamma_0)\;\; S_b^{-1}(\gamma_0)\xi_n + o_P(1)$$

where the first term on the right hand side (rhs) is present even if the true threshold parameter value γ_0 is observed and the second term on the rhs is due to the estimation of γ_0 by $\hat{\gamma}$. Just like $d_b(x,\gamma)$, both $c_{1b}(x,\gamma)$ and $c_{2b}(x,\gamma)$ also behave like a scaled univariate kernel function such that $E|c_{jb}(X_t,\gamma)|^2 = O(1/b)$ for $j = 1, 2$. This implies that even if one observes γ_0, one can only estimate the jump size β_0 at the $\sqrt{nb}$-rate.

Second, as shown in the proof of Theorem 3.3(ii),

$$\sqrt{nb^d}\left[\hat{\alpha}_b(x,\hat{\gamma}_{\pm}) - \alpha_0(x)\right] = f(x)^{-1}\frac{\sqrt{b^d}}{\sqrt{n}}\sum_{t=1}^n K_b(X_t - x)\varepsilon_t - b^{(d-1)/2}c_{\alpha,b}(x)\xi_n + o_P(1),$$

where the first term on the rhs is present even if we observe γ_0 and β_0, and the second term indicates the effect of the estimation of both γ_0 and β_0. To see the separate effects of the estimation γ_0 and β_0, we can rewrite

$$\begin{aligned} c_{\alpha,b}(x) &= f(x)^{-1}S_b^{-1}(\gamma_0)\beta_0 b\dot{c}_{0b}(x,\gamma_0)\bar{c}_{0b}(\gamma_0) - f(x)^{-1}S_b^{-1}(\gamma_0)c_{0b}(x,\gamma_0)\left[1 - \beta_0 b\dot{\theta}_{\psi_1,\pm}(\gamma_0)\Gamma_{1b,\pm}^{-1}\bar{c}_{0b}(\gamma_0)\right] \\ &\equiv c_{\alpha,b,1}(x) - c_{\alpha,b,2}(x),\ \text{say}. \end{aligned}$$

Then $b^{(d-1)/2}c_{\alpha,b,1}(x)\xi_n$ and $-b^{(d-1)/2}c_{\alpha,b,2}(x)\xi_n$ signal the effects of the estimation of γ_0 and β_0, respectively. In the special case where $d > 1$, $\Delta_{\alpha,\pm}(x;d) = 0$ and the estimation of (β,γ) does not have any asymptotic effect on the asymptotic distribution of $\hat{\alpha}_b(x,\hat{\gamma}_{\pm})$. In addition, by Assumption A3, $c_{0b}(x,\gamma_0) = 0$ if $x_1 \leq \gamma_0 - b$, $c_{0b}(x,\gamma_0) = f(x) + O(b^{\upsilon})$ if $x_1 \geq \gamma_0 - b$, and $\dot{c}_{0b}(x,\gamma_0) = 0$ if $x_1 \notin [\gamma_0 - b, \gamma_0 + b]$. With these, we further make the following two observations:

(i) If $x_1 \leq \gamma_0 - b$, $\sqrt{nb^d}\left[\hat{\alpha}_b(x,\hat{\gamma}_{\pm}) - \alpha_0(x)\right] \xrightarrow{d} N\left(0, f(x)^{-1}\sigma^2(x)\int K(u)^2du\right)$;

(ii) If $x_1 > \gamma_0 + b$, $\sqrt{nb^d}\left[\hat{\alpha}_b(x,\hat{\gamma}_{\pm}) - \alpha_0(x)\right] \xrightarrow{d} N\left(0, f(x)^{-1}\sigma^2(x)\int K(u)^2du + \bar{\Delta}_{\alpha,\pm}(x;d)\right)$, where $\bar{\Delta}_{\alpha,\pm}(x;d) = \lim_{n\to\infty}\bar{\Delta}_{\alpha,b,\pm}(x;d)$,

$$\bar{\Delta}_{\alpha,b}(x;d) = b^{(d-1)}\bar{c}_{\alpha,b,\pm}^2(x)E\{b[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)]^2\sigma^2(X_t)\}$$

and $\bar{c}_{\alpha,b,\pm}(x) = -S_b^{-1}(\gamma_0)[1 - \beta_0 b\dot{\theta}_{\psi_1,\pm}(\gamma_0)\Gamma_{1b,\pm}^{-1}\bar{c}_{0b}(\gamma_0)]$.

Third, $\hat{\beta}_b(\hat{\gamma}_{\pm})$ is not asymptotically independent of $\hat{\gamma}_{\pm}$, and it is not asymptotically independent of $\hat{\alpha}_b(x,\hat{\gamma}_{\pm})$ in the case $d = 1$. Equations (A.38) and (A.43) in Appendix A suggest both $\sqrt{nb}\left(\hat{\beta}_b(\hat{\gamma}_{\pm}) - \beta_0\right)$ and $\sqrt{n/b}(\hat{\gamma}_{\pm} - \gamma_0)$ are proportional to

$$\xi_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n b^{1/2}[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)]\varepsilon_t,$$

which explains dependence between the two. Similarly, when $d = 1$, $\sqrt{nb^d}\left[\hat{\alpha}_b\left(x, \hat{\gamma}_{\pm}\right) - \alpha_0\left(x\right)\right]$ contains a term that is linear in ξ_n, which is not asymptotically negligible. This explains the dependence between $\hat{\gamma}_{\pm}$ and $\hat{\alpha}_b\left(x, \hat{\gamma}_{\pm}\right)$. On the other hand, if $d > 1$, the linear term associated with ξ_n in the influence function of $\sqrt{nb^d}\left[\hat{\alpha}_b\left(x, \hat{\gamma}_{\pm}\right) - \alpha_0\left(x\right)\right]$ is asymptotically negligible, and then we have asymptotic independence between $\hat{\gamma}_{\pm}$ and $\hat{\alpha}_b\left(x, \hat{\gamma}_{\pm}\right)$.

Fourth, following the fourth remark after Theorem 3.2 we can also consider statistical inference for β_0 and $\alpha_0\left(x\right)$. Given the fact that γ_0 is generally unobserved and one does not know the sign of $\hat{\gamma} - \gamma_0$, we recommend the use of IID bootstrap method as in Seo and Linton (2007). We shall evaluate the finite sample performance of this bootstrap method via simulations.

4 Finite sample performance

Here we consider the finite sample performance of our estimators/tests via Monte Carlo simulations. Given the general nature of our estimator, we consider several different scenarios: (1) cross-sectional data where the threshold variable is also a regressor included (2) cross-sectional data where the threshold variable is excluded (3) time-series data where the threshold variable is included (4) time-series data where the threshold variable is excluded.

Our performance criteria for evaluating our estimator of $\alpha(x)$ is weighted average squared error (WASE),

$$WASE(\widehat{\alpha}(x)) = n^{-1}\sum_{t=1}^{n}\left(\widehat{\alpha}(x_t) - \alpha(x_t)\right)^2 1\left\{\left|\frac{x_t - \bar{x}}{\sigma_x}\right| \leq 2\right\}, \tag{4.1}$$

where $\widehat{\alpha}(x)$ is our Nadaraya-Watson estimator of the unknown function. $WASE$ is evaluated at the sample points for each simulation. For our threshold effect, β, and our unknown threshold, γ, we report bias and mean squared error across the simulations. For each DGP we consider the case of both known and unknown threshold parameter γ. Unless otherwise stated, we use sample sizes of $n = 100$, 200 and 400 with 1000 replications per experiment. For all simulations, we use a second order Epanechnikov kernel with rule-of-thumb bandwidth, $b = 2.345 \cdot \widehat{\sigma}_x n^{-0.25}$, where $\widehat{\sigma}_x$ is the sample standard deviation of the covariate. We use an undersmoothed bandwidth given the remarks pertaining to Assumption 4.

Inference about the threshold parameter γ can be examined using the large sample results in Theorem 3.2. Although our semiparametric threshold estimator possesses a limiting normal distribution, its variance is somewhat complicated. As an alternative to direct estimation of this variance, we can use the bootstrap. However, as laid out in exceeding detail by Yu (2012), standard bootstrap approaches will not work. The reason is that the threshold parameter represents a boundary and common bootstrap mechanisms are known to be invalid when a boundary exists. Further, as evidenced in both Yu (2012) and Seo and Linton (2007), both percentile t and pivotal bootstrap approaches do not produce correct coverage of the threshold parameter (see also Seijo and Sen; 2011). Footnote 3 in Seo and Linton (2007) indicates that the bootstrap may be inconsistent in their simulations, but no theoretical analysis is conducted on this point.

We will demonstrate how well our estimator works using the smooth bootstrap for our threshold estimator for all of the DGPs. To detail how the smooth bootstrap works we follow the insight of Silverman (1986) and sample from $\widehat{f}(x, q)$, the kernel density estimator of our covariates. In the case where $x = q$, we construct resamples from the univariate kernel density estimator. However, the density does not actually need to be constructed. Rather, smoothed bootstrap observations can be constructed as

$$w_t^* = (x_t^*, q_t^*) = (1 + b^2/\widehat{\sigma}^2)(w_{(t)} + b\epsilon_t)$$

where $w_{(t)}$ is sampled uniformly with replacement from the original data, b is a bandwidth vector, $\widehat{\sigma}^2$ is the vector of estimated variances for the data and ϵ_t is a random draw from a multivariate normal with mean 0 and variance $\widehat{\Sigma}$, the sample covariance of w_t.

4.1 Cross-section where threshold variable is a regressor

We have a semiparametric threshold model error

$$g(X_t) = \alpha(x_t) + \beta \cdot 1\left\{x_t > \gamma\right\}.$$

Here we investigate seven function specifications for $\alpha(x)$:

CSB DGP 1 $\alpha(x) = 0.8 + 0.7x$;

CSB DGP 2 $\alpha(x) = 2 + 1.8\sin(1.5x)$;

CSB DGP 3 $\alpha(x) = 2.75\frac{e^{-3x}}{1+e^{-3x}} - 1$;

CSB DGP 4 $\alpha(x) = 0.7x + 1.4e^{-16x^2}$;

CSB DGP 5 $\alpha(x) = 1.05(\cos(\pi x) + \sin(\pi x) + \log(7/3 + x/2))$;

CSB DGP 6 $\alpha(x) = 2(x^4 - 0.1x^3 - 4.64x^2 + 1.324x + 0.408)/17$;

CSB DGP 7 $\alpha(x) = 0.2 + 0.3x - 0.41x^2$.

Our parameters come from $\gamma \in \{-1, 0, 1\}$ and $\beta \in \{1, 1.5, 2\}$. We divide the signal component $g(X_t)$ for each DGP by its standard deviation, $\sigma_{g(X_t)}$ to control the signal-to-noise ratio, generating our dependent regressor as $Y_t = g(X_t)/\sigma_{g(X_t)} + \varepsilon_t$ where $x \sim \mathcal{U}[-3, 3]$. Lastly, we take $\varepsilon_t \sim N(0, \sigma^2)$, with $\sigma \in \{0.32, 0.58, 0.82\}$ which yields signal-to-noise ratios of 0.9, 0.75 and 0.6, respectively.

For brevity we only report the results for DGP, for $\gamma = -1, 0$. These results appear in Tables 1 to 3. Several key features emerge: as n increases the bias of both $\widehat{\beta}$ and $\widehat{\gamma}$ decrease, the WASE for all three estimators decrease as the sample size increases, with the rate of decrease for $\widehat{\gamma}$ faster, as expected, than $\widehat{\beta}$ and $\widehat{\alpha}(x)$. We also notice that as the signal-to-noise ratio increases the performance of our estimator improves for all sample sizes. This is expected as the threshold location is easier to identify with less noise.

Results for DGPs 2-7 are similar. We provide simulation results for the remaining DGPS, using a signal-to-noise ratio of 0.75 in Appendix C.

Table 1: Simulation Performance of Semiparametric Threshold Estimator, DGP 1, signal to noise ratio=0.9, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-0.093	0.283	0.200	0.636	0.025
$n = 200$	-0.028	0.109	0.071	0.204	0.014
$n = 400$	-0.004	0.043	0.002	0.005	0.008
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-0.039	0.326	0.052	0.162	0.028
$n = 200$	-0.024	0.106	0.007	0.024	0.015
$n = 400$	-0.010	0.042	-0.001	0.000	0.009
	$\beta = 2, \gamma = -1$				
$n = 100$	-0.047	0.321	0.016	0.061	0.029
$n = 200$	-0.008	0.106	-0.004	0.001	0.017
$n = 400$	-0.012	0.052	0.000	0.000	0.009
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.086	0.306	0.036	0.406	0.026
$n = 200$	-0.035	0.155	0.025	0.138	0.014
$n = 400$	-0.006	0.036	0.006	0.011	0.008
	$\beta = 1.5, \gamma = 0$				
$n = 100$	-0.103	0.378	0.000	0.189	0.028
$n = 200$	-0.015	0.112	-0.008	0.011	0.015
$n = 400$	-0.003	0.049	0.000	0.000	0.009
	$\beta = 2, \gamma = 0$				
$n = 100$	-0.070	0.351	-0.006	0.067	0.031
$n = 200$	-0.009	0.125	-0.004	0.005	0.017
$n = 400$	0.003	0.054	0.000	0.000	0.010

5 Empirical examples

5.1 Threshold: Fiscal cliff

The importance (or lack thereof) of public debt on economic development is a controversial topic within academic and policy debates. This issue is all the more important given the recent global downturn, spanning both the developed and developing worlds. Existing studies are compromised by the fact that their focus is on developed economies, or focus exclusively on a small set of seemingly similar developing

Table 2: Simulation Performance of Semiparametric Threshold Estimator, DGP 1, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-0.250	1.273	0.522	1.505	0.070
$n = 200$	-0.124	0.663	0.327	0.966	0.036
$n = 400$	-0.065	0.319	0.142	0.405	0.017
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-0.236	1.593	0.306	0.924	0.078
$n = 200$	-0.155	0.807	0.144	0.406	0.035
$n = 400$	-0.026	0.236	0.026	0.078	0.019
	$\beta = 2, \gamma = -1$				
$n = 100$	-0.205	1.763	0.220	0.658	0.076
$n = 200$	-0.085	0.647	0.075	0.188	0.038
$n = 400$	-0.018	0.215	0.006	0.019	0.021
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.263	1.405	0.035	1.045	0.072
$n = 200$	-0.225	0.881	-0.026	0.667	0.038
$n = 400$	-0.127	0.395	0.010	0.330	0.018
	$\beta = 1.5, \gamma = 0$				
$n = 100$	-0.284	1.758	-0.056	0.640	0.080
$n = 200$	-0.154	0.860	0.019	0.311	0.037
$n = 400$	-0.019	0.279	0.004	0.070	0.020
	$\beta = 2, \gamma = 0$				
$n = 100$	-0.314	2.098	0.006	0.504	0.087
$n = 200$	-0.118	0.913	0.008	0.161	0.044
$n = 400$	-0.017	0.302	0.003	0.020	0.023

countries. In general, existing estimates suggest an optimal public debt ratio of 30-70% of GDP.

However, these numbers must be viewed with caution when considering policy prescriptions for developing countries. A positive view of public debt exists that promotes the use of public debt as an instrument for both financial and monetary systems within low income countries as well as for overall development of political institutions. This is illustrated with the recent experiences of China, India and Chile, all of whom have been able to maintain low levels of external indebtedness and avoided major financial and fiscal crises. This stems from the ability of domestic debt to contribute to macroeconomic stability through low inflation as well as private savings accumulation and investment.

Table 3: Simulation Performance of Semiparametric Threshold Estimator, DGP 1, signal to noise ratio=0.6, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	−0.408	2.748	0.750	1.990	0.132
$n = 200$	−0.341	1.664	0.585	1.611	0.071
$n = 400$	−0.235	0.953	0.342	0.985	0.037
	$\beta = 1.5, \gamma = -1$				
$n = 100$	−0.380	3.296	0.563	1.538	0.160
$n = 200$	−0.375	2.173	0.391	1.063	0.083
$n = 400$	−0.170	0.952	0.175	0.512	0.037
	$\beta = 2, \gamma = -1$				
$n = 100$	−0.509	4.380	0.455	1.230	0.192
$n = 200$	−0.315	2.289	0.255	0.699	0.084
$n = 400$	−0.076	0.809	0.065	0.204	0.040
	$\beta = 1, \gamma = 0$				
$n = 100$	−0.371	2.847	−0.032	1.225	0.133
$n = 200$	−0.396	1.811	0.021	1.128	0.077
$n = 400$	−0.228	0.959	0.021	0.668	0.039
	$\beta = 1.5, \gamma = 0$				
$n = 100$	−0.513	3.977	−0.027	1.035	0.179
$n = 200$	−0.415	2.254	−0.030	0.749	0.089
$n = 400$	−0.201	1.033	−0.011	0.353	0.039
	$\beta = 2, \gamma = 0$				
$n = 100$	−0.685	5.149	0.044	0.874	0.206
$n = 200$	−0.341	2.588	−0.007	0.500	0.095
$n = 400$	−0.162	1.061	0.007	0.193	0.045

The objective of this example is twofold. We exploit a recently developed domestic debt database published by the IMF with excellent time and individual coverage to analyze the role that public debt has on economic growth. This in and of itself is a contribution. However, we augment these results, stemming from popular threshold regression models, by employing our semiparametric threshold model.

Finally, not to beat a dead horse, but it may be useful to compare our estimated threshold to that of Reinhart and Rogoff (2010) who infamously argued that "whereas the link between growth and debt seems relatively weak at 'normal' debt levels, median growth rates for countries with public debt over roughly 90 percent of GDP are about one percent lower than otherwise; average (mean) growth rates are several percent lower."

5.2 Data

Our data for the standard Solow variables comes directly from Henderson, Papageoriou and Parmeter (2013). The data for public debt comes directly from Abbas and Christensen (2007). We will explain each data set briefly.

The data in Henderson, Papageoriou and Parmeter (2013) is partially taken from Durlauf, Kourtellos and Tan (2008). In the latter paper, the data for per capita real GDP and the average growth rate of the working age population are taken from the Penn World Tables, Version 6.1. The data for investment is obtained via the capital per worker variable in Caselli (2005). The data for education, which is measured as the average percentage of the working age population (population between the age of 15 and 64) in secondary education is taken from Barro and Lee (2000).

The public debt data, defined as "commercial banks' gross claims on the central government *plus* central bank liquidity paper," comes directly from Abbas and Christensen (2007) which is primarily based on Abbas (2007a) who obtains his data from the International Financial Statistics monetary survey. The debt data is scaled by the corresponding GDP data.

The combination of these two data sets results in an unbalanced panel of 90 countries over the period 1965-1995. This results in a total of 534 observations for our sample. The entire data set used here is available from the authors upon request.

5.3 Results

We construct a balanced panel of 65 countries over the period 1970-1995, in five year intervals constituting 390 observations. We estimate two distinct models to begin, first a generic human capital augmented Solow growth model using public debt as a threshold, and second, we include public debt directly into our growth model and still look for a threshold with respect to public debt.

5.3.1 Parametric

We first use the test of a threshold proposed by Hansen (1996) to determine if a threshold exists in public debt. Whether public debt is included as a covariate directly or not, we obtain a bootstrap p-value of 0. If the parametric model is correctly specified, this provides evidence of a threshold in public debt.

Table 4 presents estimates via the estimator in Hansen (2000) for our two models. We list the coefficient and it's corresponding heteroskedasticity-robust standard error as well as the threshold estimate from each model with the corresponding upper and lower decile bootstrapped estimates. We see that public debt has a statistically significant effect. Moreover, the estimation of the threshold is more accurate in the model including public debt as a threshold. In model (1) 371 out of 390 observations (*approx* 95%) fall within the confidence band for the public debt threshold. This places doubt on the classification of countries into groups based on the public debt threshold. However, once public debt is included as a regressor, only 127 out of 390 observations (≈ 33%) fall within the confidence band, providing a much stronger ability to segment countries.

Table 4: Hansen (2000) estimates for models both without (1) and with (2) public debt included as a regressor. The left-hand-side variable is logarithmic growth over the previous five year interval. Heteroscedasticity-robust standard errors are reported in parentheses beneath each estimate. The threshold parameter is listed along with the 10th and 90th percentiles in brackets.

	(1)	(2)
Constant	-0.1294	-0.1254
	(0.0293)	(0.0299)
GDP Lag	-0.0017	-0.0024
	(0.0028)	(0.0028)
Investment/GDP	0.0178	0.0176
	(0.0041)	(0.0041)
Pop Growth	-0.0435	-0.0447
	(0.0117)	(0.0121)
School	-0.0013	-0.0011
	(0.0008)	(0.0008)
Debt/GDP	—	-0.0000178
		(0.0000075)
Threshold Estimate	0.9099	0.5629
	[0.0927,2.1801]	[0.4853,0.9318]

We also considered separating observations into the corresponding regimes based on the estimated thresholds in Table 4. We see that the impact of public debt levels on growth is roughly 15 times larger (in magnitude) in Regime 1 than Regime 2, suggesting that lower levels of public debt help growth more than high levels of public debt.

5.3.2 Semiparametric

Here we take the two models estimated above and run semiparametric versions of them. In Table 5, we give the median gradient estimate for each regressor (roughly comparable to the slope coefficient estimates in Table 4) for each semiparametric model along with the corresponding bootstrapped standard errors. As before, we report the threshold estimate from each model with the corresponding upper and lower decile bootstrapped estimates.

We can see that the median estimates are similar to those from the parametric model. This is a common phenomenon. That being said, there is significant variation in the point estimates from the nonparametric model and these can often lead to major differences across groups of countries (e.g., see Henderson, Papageorgiou and Parmeter 2012,2013).

Table 5: Semiparametric estimates for models both without (1) and with (2) public debt included as a regressor. The left-hand-side variable is logarithmic growth over the previous five year interval. (Smooth) bootstrapped standard errors are reported in parentheses beneath each estimate. The threshold parameter is listed along with the 10th and 90th percentiles in brackets.

	(1)	(2)
GDP Lag	-0.0044	-0.0049
	(0.0077)	(0.0051)
Investment/GDP	0.0129	0.0124
	(0.0129)	(0.0090)
Pop Growth	-0.0164	-0.0118
	(0.0514)	(0.0329)
School	-0.0006	-0.0006
	(0.0026)	(0.0017)
Debt/GDP	—	-0.0017
		(0.0062)
Threshold Estimate	0.8321	0.6848
	[0.4639,0.8689]	[0.5743,0.7216]

The threshold estimates are perhaps equally interesting. The table shows that for the case where debt is not a regressor that the threshold point estimate is 0.8321 which is comparable to the parametric result for the same model (0.9099). Similarly, for the case where debt is a regressor we get a smaller threshold estimate (0.6848), but this is now larger than the corresponding parametric estimate (0.5629). The confidence bounds for the estimates overlap between estimators, but we can see that the nonparametric estimates are obtained with less variability.

The placement of the threshold is important, but the impact of being on one side or the other (β) is also relevant. When debt is not included as a regressor we get an estimate of β equal to -0.0051. This is the expected sign, and when taken literally, implies that increasing the debt ratio above the threshold leads to a drop in the growth rate of GDP. On the other hand, when debt is included as a regressor, $\widehat{\beta} = 0.01142$. That being said, each of these estimates are near zero in a statistical sense. Again, note that the method looks for the most likely break point. It is feasible that there is no (single hard) break (in terms of debt to DGP ratio) in this series.

5.4 Weak dependence: Asymmetric time series

The second example we consider is a univariate time series. Here we model perhaps the most studied (univariate) time series in macroeconomics: U.S. GNP. Beginning with Hamilton (1989), there is a long history modeling U.S. GNP non-linearly (e.g., Beaudry and Koop, 1993; Potter, 1995; Terasvirta and

Anderson, 1992). The term nonlinear in this literature is often different from what we have discussed earlier in the paper. In the example to which we compare our paper, nonlinear implies a linear model with parameters which vary based on the sign of one of the regressors. Our plan is to allow the function to vary based on a threshold, but will relax the linearity assumption on either side of the threshold by using semiparametric regression.

5.4.1 Data

Our data come directly from Potter (1995) and we only explain them briefly. Real U.S. GNP, taken from the Citibase data bank, is seasonally adjusted and measured quarterly from the first quarter of 1947 to the fourth quarter of 1994. Following Potter (1995), to obtain the growth rates, these values are measured in logged first differences and multiplied by 100.

5.4.2 Results

We compare our estimator to that given in Table III of Potter (1995). In that table, he presents a fifth order autoregressive model (lags at 1, 2 and 5 quarters) with a threshold based on the second lag at zero ($Y_{t-2} \geq 0$) – expansions versus contractions. Our semiparametric threshold alternative model is given as

$$Y_t = \alpha\left(Y_{t-1}, Y_{t-2}, Y_{t-5}\right) + \beta \cdot 1\left\{Y_{t-2} \geq \gamma\right\} + \varepsilon_t,$$

where we do not assume that $\gamma = 0$. Here we both wish to compare both the assumption that the break is at zero and the performance of the kernel versus the parametric estimator.

Figure 1 gives the time series along with the fit from the semiparametric model and the estimated threshold. The vertical line that represents the estimated threshold is equal to 0.0076.The confidence intervals include the value zero. It is interesting to note that the estimated threshold is nearly identical to the mean of the time series (0.0077). Given that this is quarterly data, that represents roughly a 3% annual growth rate. Taking these results literally implies that the series behaves differently above and below it's long run average.

As for the comparison between the parametric and semiparametric models, we find a much smaller standard error of the regression (0.0044 versus 0.95597). That being said, we should be cautious about over-fitting with a semiparametric alternative. We found similar improvements over the parametric model in terms of R^2, AIC and SBC (using the definitions in Gao, 2007). We also considered the case where we restricted our semiparametric threshold value to be at zero ($\gamma = 0$) and found similar improvements in terms of in-sample fit ($\widehat{\sigma} = 0.0714$)

5.5 Regression discontinuity: U.S. elections

Although formally elected, the U.S. House of Representatives appears aristocratic. In 2012, 90 percent of House members who ran for re-election were successful. "Incumbency advantage" is well known and is defined as the "overall causal impact of being the current incumbent party in a district on the votes obtained in the district's election" (Lee, 2008).

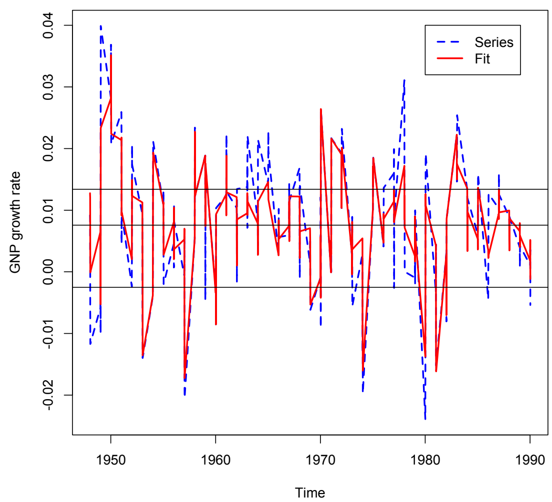

Figure 1: Time series plot of the quarterly growth rate of U.S. GNP along with the fit from the semiparametric model and estimated threshold parameter (with corresponding 95 percent confidence bounds)

In his well cited article, Lee (2008) establishes the conditions under which a regression discontinuity analysis can be seen as credible as those from a randomized experiment. In his application, he considers re-election in the U.S. House of Representatives. He argues that the relationships "exhibit important non-linearities" and that "a linear regression specification would hence lead to misleading inferences." Taking his lead, we use the identical data from his paper to re-examine a subset of his results in order to show how our semiparametric estimators work in an RDD framework.

The discontinuity point here is well known (being elected in the previous term), but we will assume it to be unknown to show how our estimators can correctly estimate the break point.

5.5.1 Data

Our data come directly from Lee (2008) and we only explain them briefly. The data are based on both the (ICPSR study 7757) "Candidate and Constituency Statistics of Elections in the United States, 1788-1990)" study (ICPSR, 1995) and United States House of Representatives Office of the Clerk's Web Page for the years 1992-1998. Lee (2008) checked for internal consistencies and uses the sample period 1946–1998.[2] The sample consists of 9674 Democrat[3] observations over the sample period. Although in nearly all cases the strongest opponent was a Republican, third party candidates do exist and hence winning an election does not require 50% of the vote. Hence, in order to determine whether or not the Democratic candidates wins the election, the explanatory variable of interest (Democratic vote share margin of victory in period t) has a known threshold at zero (positive values lead to winning an election and negative values to losing an election). We consider two of the left-hand-side variables in Lee (2008): (1) winning the election in period $t+1$ and (2) candidacy in period $t+1$ (where $t+1$ refers to the next election cycle – every 2 years). It is argued that winning an election (even by a narrow margin) in period t leads to much higher values of the left-hand-side variables.

5.5.2 Results

Here we give our take on two results in Lee (2008). The first is analogous to his Figure 2(a). Our Figure 2 (with confidence bounds excluded for clarityj) shows the RD estimate of incumbency advantage. The horizontal axis gives the difference in the Democratic vote share and that of the strongest opponent in period t. Values greater than zero represent winning the election and values less than zero represent losing the election. The vertical axis gives the probability of running and being elected in the next election cycle (period $t+1$).

The known break point here is zero and our estimator correctly gives this value ($\widehat{\gamma}=0$).[4] This gives us confidence in our estimator in an empirical application. It is obvious that there is a big difference from narrowly losing to narrowly winning an election in period t on period $t+1$'s outcome. The estimated

[2] Several points had to be imputed and the details can be found in Appendix A of Lee (2008, pp. 693).

[3] Lee (2008) only considers Democrat candidates as in nearly all elections (in a two-party system) the opponent is a Republican and hence a winning Democrat produces a losing Republican (and vice versa). Given the relatively small numbers of third party candidates, he argues that studying Republican candidates will give 'mirror image' results.

[4] Note that we included the value of 0 over our grid of possible break points (as well as many points near 0).

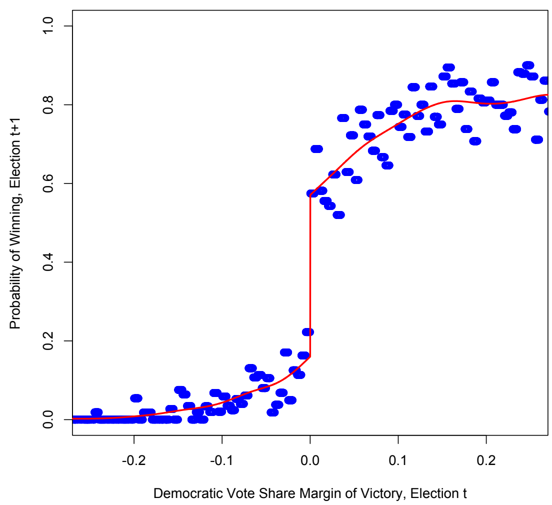

Figure 2: Probability of Democrat running and winning election in period t+1 versus margin of victory in period t (values on the horizontal axis greater than zero imply winning an election in period t)

causal effect ($\widehat{\beta}$) is 0.4107, which is slightly smaller than the causal effect reported in Lee (2008) of approximately 0.45 in probability.

It is worth pointing out the nonlinear relationship both before and after the break. Our fit roughly resembles the fit of the function in Lee (2008). The probability increases at an increasing rate prior to zero and then increases at a decreasing rate past zero. It would likely be difficult to reject his estimates in a formal test.

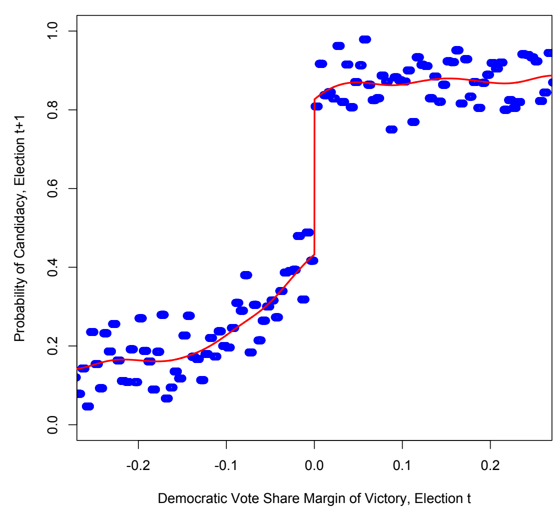

Figure 3: Probability of candidacy in year t+1 versus margin of victory in period t (values on the horizontal axis greater than zero imply winning an election in period t)

Our second comparison (Figure 3), analogous to his Figure 3(a), shows the probability that the Democrat remains the nominee for the party in period $t+1$ given the election result in period t. Note again that we correctly estimate the break point of winning the election ($\widehat{\gamma} = 0$). The RD estimate ($\widehat{\beta} = 0.3941$) is nearly as large as that in the previous figure (this result is nearly identical to that in Lee, 2008). This shows that a narrow margin of victory makes a huge difference on whether or not the candidate decides to run for re-election in the next cycle.

Even though this, and the previous two examples are relatively simple, they demonstrate that our proposed estimator can handle a range of different scenarios and provide meaningful insights. Each of these, as well as other applications, deserve a more rigorous treatment in future research.

6 Conclusion

We have detailed a super consistent estimator for an unknown threshold in the context of a nonparametric regression model. This estimator used three steps to recover the model primitives. Relying on semiparametric M-estimation we detailed the large sample properties of our proposed estimators for the unknown threshold, the size of the jump of the function at the threshold and the unknown conditional mean.

A series of Monte Carlo simulations and several empirical examples highlighted the practical merits of the method while our theoretical results extended the seminal contributions of Pakes and Porter (1986) and Chen, Linton and Van Keilgom (2003) to allow for semiparametric extremum estimation when the objective function is flat.

Appendix

A Proof of the Results in Section 3

We first prove some lemmas that are used in the proof of the main results in Section 3.

Lemma A.1 *Suppose that Assumptions A1, A2(i), A3, and A.4 hold. Then* $\sup_{x\in\mathcal{X}_0}\sup_{(\beta,\gamma)\in\mathcal{B}\times\Gamma}|\hat{\alpha}_b(x;\beta,\gamma)-\alpha_{0,b}(x;\beta,\gamma)|=O_P(b^{\upsilon}+\nu_n)$ *where* $\alpha_{0,b}(x;\beta,\gamma)=\alpha_0(x)+\delta_b(x;\beta,\gamma)$, $\delta_b(x;\beta,\gamma)\equiv f(x)^{-1}E\{K_b(X_t-x)[\beta_0D_t(\gamma_0)-\beta D_t(\gamma)]\}$, *and* $\nu_n\equiv(n^{-1}b^{-d}\log n)^{1/2}$.

Proof. Noting that $Y_t=\alpha_0(X_t)+\beta_0D_t(\gamma_0)+\varepsilon_t=\alpha_0(x)+[\alpha_0(X_t)-\alpha_0(x)]+\beta_0D_t(\gamma_0)+\varepsilon_t$, by (2.15) we have

$$\begin{aligned}\hat{\alpha}_b(x;\beta,\gamma) &= \hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)[Y_t-\beta D_t(\gamma)]\\ &= \alpha_0(x)+\hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)\varepsilon_t\\ &\quad +\hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)[\alpha_0(X_t)-\alpha_0(x)]\\ &\quad +\hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)[\beta_0D_t(\gamma_0)-\beta D_t(\gamma)]\\ &\equiv \alpha_0(x)+V_{nb}(x)+B_{nb}(x)+R_{nb}(x;\beta,\gamma),\text{ say,}\end{aligned}\tag{A.1}$$

where $B_{nb}(x)$ and $V_{nb}(x)$ denote the standard asymptotic bias and variance terms of $\hat{\alpha}_b(x;\beta,\gamma)$, and the remainder term $R_{nb}(x;\beta,\gamma)$ is new. Following the arguments used in Masry (1996) and Hansen (2008), we can easily show that

$$\sup_{x\in\mathcal{X}_0}\left|\hat{f}_b(x)-f(x)\right|=O_P(b^{\upsilon}+\nu_n),\quad \sup_{x\in\mathcal{X}_0}|V_{nb}(x)|=O_P(\nu_n),\quad \sup_{x\in\mathcal{X}_0}|B_{nb}(x)|=O_P(b^{\upsilon}),\tag{A.2}$$

and

$$\sup_{x\in\mathcal{X}_0}\sup_{(\beta,\gamma)\in\mathcal{B}\times\Gamma}\left|R_{nb}(x;\beta,\gamma)-\delta_b(x;\beta,\gamma)\right|=O_P\left(b^{\upsilon}+\nu_n\right). \tag{A.3}$$

Combining (A.1)-(A.3) yields the conclusion. ■

Lemma A.2 *Suppose that Assumptions A1, A2(i), A3, and A4 hold. Then*

(i) $\hat{\beta}_b(\gamma)-\beta_{0,b}(\gamma)=S_b^{-1}(\gamma)\frac{1}{n}\sum_{t=1}^n\left[c_{1b}(X_t,\gamma)-c_{2b}(X_t,\gamma)\right]\varepsilon_t+\beta_0 S_b(\gamma)^{-1}\frac{3}{n}\sum_{t=1}^n\{\psi_{1b}(X_t;\gamma)-E\left[\psi_{1b}(X_t;\gamma)\right]\}+O_P(b^{\upsilon})+o_P\left((nb)^{-1/2}\right)$ *uniformly in* $\gamma\in\Gamma$,

(ii) $\sup_{\gamma\in\Gamma}\left\|\hat{\beta}_b(\gamma)-\beta_{0,b}(\gamma)\right\|=O_P\left(b^{\upsilon}+(nb/\log n)^{-1/2}\right)$,

where $\beta_{0,b}(\gamma)=\beta_0+\beta_0 c_{d_b}(\gamma)$ *with* $c_{d_b}(\gamma)$ *defined in (3.2),* $S_b(\gamma)$ *is defined below (3.7),* $c_{1b}(x,\gamma)$ *and* $c_{2b}(x,\gamma)$ *are defined in (3.6), and* $\psi_{1b}(\cdot;\gamma)$ *is defined in (A.12).*

Proof. (i) Noting that $Y_t=\alpha_0(X_t)+\beta_0 D_t(\gamma_0)+\varepsilon_t$, we have

$$\tilde{Y}_t=n^{-1}\sum_{s=1}^n K_b(X_s-X_t)(Y_s-Y_t)=\beta_0\tilde{D}_t(\gamma)+\tilde{\varepsilon}_t+\tilde{\alpha}_0(X_t)+\beta_0[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)],$$

where $\tilde{D}_t(\gamma)=n^{-1}\sum_{s=1}^n K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]$, $\tilde{\varepsilon}_t=n^{-1}\sum_{s=1}^n K_b(X_s-X_t)(\varepsilon_s-\varepsilon_t)$, and $\tilde{\alpha}_0(X_t)=n^{-1}\sum_{s=1}^n K_b(X_s-X_t)\left[\alpha_0(X_s)-\alpha_0(X_t)\right]$. It follows from (2.17) that

$$\begin{aligned}\hat{\beta}_b(\gamma) &= S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^n\tilde{D}_t(\gamma)\tilde{Y}_t\\ &= \beta_0+S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^n\tilde{D}_t(\gamma)\tilde{\varepsilon}_t+S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^n\tilde{D}_t(\gamma)\tilde{\alpha}_0(X_t)\\ &\quad+\beta_0 S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^n\tilde{D}_t(\gamma)[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)]\\ &\equiv \beta_0+v_{nb}(\gamma)+b_{nb}(\gamma)+r_{nb}(\gamma),\ \text{say},\end{aligned} \tag{A.4}$$

where $S_{nb}(\gamma)=\frac{1}{nb}\sum_{t=1}^n\tilde{D}_t(\gamma)^2$.

We first study $S_{nb}(\gamma)$. Observe that

$$\begin{aligned}S_{nb}(\gamma) &= \frac{1}{nb}\sum_{t=1}^n\left\{n^{-1}\sum_{s\neq t}^n\{K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]-b\cdot d_b(X_t,\gamma)\}+b\cdot d_b(X_t,\gamma)\right\}^2\\ &= \frac{b}{n}\sum_{t=1}^n d_b(X_t,\gamma)^2+\frac{2}{n}\sum_{t=1}^n d_b(X_t,\gamma)\,n^{-1}\sum_{s\neq t}^n\{K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]-b\cdot d_b(X_t,\gamma)\}\\ &\quad+\frac{1}{nb}\sum_{t=1}^n\left\{n^{-1}\sum_{s\neq t}^n\{K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]-b\cdot d_b(X_t,\gamma)\}\right\}^2\\ &\equiv S_{nb,1}(\gamma)+2S_{nb,2}(\gamma)+S_{nb,3}(\gamma),\ \text{say}.\end{aligned}$$

Using arguments as used in Masry (1996) or Hansen (2008), we can readily show that

$$\sup_{\gamma\in\Gamma}\left|S_{nb,1}(\gamma)-S_b(\gamma)\right|=O_P(n^{-1/2}b^{-1/2}(\log n)^{1/2}).$$

Let $\varphi^0(X_t,X_s;\gamma)=d_b(X_t,\gamma)\{K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]-b\cdot d_b(X_t,\gamma)\}$ and $\varphi(X_t,X_s;\gamma)=[\varphi^0(X_t,X_s;\gamma)+\varphi^0(X_s,X_t;\gamma)]/2$. Then

$$S_{nb,2}(\gamma)=\frac{n-1}{n}\frac{2}{n(n-1)}\sum_{1\leq s<t\leq n}\varphi(X_t,X_s;\gamma).$$

Let $\varphi_1(\cdot) = E[\varphi(\cdot, X_t; \gamma)]$ and $\varphi_2(a_1, a_2; \gamma) = \varphi(a_1, a_2; \gamma) - \varphi_1(a_1; \gamma) - \varphi_1(a_2; \gamma)$. By construction, $E[\varphi_1(X_t)] = 0$ and $E_{X_s} E_{X_t}[\varphi_2(X_t, X_s; \gamma)] = 0$, where E_{X_t} denotes expectation with respect to X_t. By Hoeffding decomposition (e.g., Lee, 1990, p.26),

$$S_{nb,2}(\gamma) = \frac{n-1}{n} \left\{ \frac{1}{n} \sum_{t=1}^{n} \varphi_1(X_t; \gamma) + \frac{2}{n(n-1)} \sum_{1 \le s < t \le n} \varphi_2(X_t, X_s; \gamma) \right\}.$$

It is standard to show that $\sup_{\gamma \in \Gamma} \left| \frac{1}{n} \sum_{t=1}^{n} \varphi_1(X_t; \gamma) \right| = O_P(n^{-1/2} b^{-1/2} (\log n)^{1/2})$. Noting that the second term in the above curly bracket is a second order degenerate U-statistic, we can modify the proof of (A.10) in Gozalo and Linton (2001) and show that

$$\sup_{\gamma \in \Gamma} \left| \frac{2}{n(n-1)} \sum_{1 \le s < t \le n} \varphi_2(X_t, X_s; \gamma) \right| = O_P\left(n^{-1} b^{-1/2} \log n\right).$$

Consequently, $\sup_{\gamma \in \Gamma} |S_{nb,2}(\gamma)| = O_P(n^{-1/2} b^{-1/2} (\log n)^{1/2})$. Similarly, we can show that $\sup_{\gamma \in \Gamma} |S_{nb,3}(\gamma)| = O_P(n^{-1/2} b^{-1/2} (\log n)^{1/2})$. It follows that

$$\sup_{\gamma \in \Gamma} |S_{nb,1}(\gamma) - S_b(\gamma)| = O_P(n^{-1/2} b^{-1/2} (\log n)^{1/2}). \tag{A.5}$$

Next, let $\tilde{v}_{nb}(\gamma) \equiv S_{nb}(\gamma) v_{nb}(\gamma)$. Then

$$\begin{aligned}
\tilde{v}_{nb}(\gamma) &= \frac{1}{nb} \sum_{t=1}^{n} \left\{ n^{-1} \sum_{s=1}^{n} K_b(X_s - X_t) [D_s(\gamma) - D_t(\gamma)] \right\} \left\{ n^{-1} \sum_{r=1}^{n} K_b(X_r - X_t)(\varepsilon_r - \varepsilon_t) \right\} \\
&= \frac{1}{n^3 b} \sum_{1 \le t \ne s \ne r \le n} K_b(X_s - X_t) K_b(X_r - X_t) [D_s(\gamma) - D_t(\gamma)] (\varepsilon_r - \varepsilon_t) \\
&\quad + \frac{1}{n^3 b} \sum_{1 \le t \ne s \le n} [K_b(X_s - X_t)]^2 [D_s(\gamma) - D_t(\gamma)] (\varepsilon_s - \varepsilon_t) \\
&\equiv \tilde{v}_{nb,1}(\gamma) + \tilde{v}_{nb,2}(\gamma), \text{ say.}
\end{aligned} \tag{A.6}$$

For $\tilde{v}_{nb,2}(\gamma)$, it suffices to use the rough bound. Noting that $|D_s(\gamma) - D_t(\gamma)| \le 1\{|\gamma - X_{1t}| \le |X_{1s} - X_{1t}|\}$, we can readily show that

$$E \sup_{\gamma \in \Gamma} |\tilde{v}_{nb,2}(\gamma)| \le \frac{2}{n^3 b} \sum_{1 \le t \ne s \le n} E\left\{ [K_b(X_s - X_t)]^2 \sup_{\gamma \in \Gamma} 1\{|\gamma - X_{1t}| \le |X_{1s} - X_{1t}|\} |\varepsilon_t| \right\} = O\left(n^{-1} b^{-d}\right).$$

By Markov inequality and Assumption A4,

$$\sup_{\gamma \in \Gamma} |\tilde{v}_{nb,2}(\gamma)| = O_P\left(n^{-1} b^{-d}\right) = o_P((nb)^{-1/2}). \tag{A.7}$$

To bound $\tilde{v}_{nb,1}(\gamma)$, let $\phi^0(\xi_t, \xi_s, \xi_r; \gamma) \equiv \frac{1}{b} K_b(X_s - X_t) K_b(X_r - X_t) [D_s(\gamma) - D_t(\gamma)] (\varepsilon_r - \varepsilon_t)$. Define its symmetric version: $\phi(\xi_t, \xi_s, \xi_r; \gamma) \equiv [\phi^0(\xi_t, \xi_s, \xi_r; \gamma) + \phi^0(\xi_t, \xi_r, \xi_s; \gamma) + \phi^0(\xi_r, \xi_t, \xi_s; \gamma) + \phi^0(\xi_r, \xi_s, \xi_t; \gamma) + \phi^0(\xi_s, \xi_t, \xi_r; \gamma) + \phi^0(\xi_s, \xi_r, \xi_t; \gamma)]/6$. Then

$$\tilde{v}_{nb,1}(\gamma) = \frac{1}{n^3} \sum_{1 \le t \ne s \ne r \le n} \phi_0(\xi_t, \xi_s, \xi_r; \gamma) = \frac{(n-1)(n-2)}{n^2} \bar{v}_{nb,1}(\gamma), \tag{A.8}$$

where $\bar{v}_{nb,1}(\gamma) = \frac{6}{n(n-1)(n-2)} \sum_{1 \le t < s < r \le n} \phi(\xi_t, \xi_s, \xi_r; \gamma)$ is a third-order U-statistic. Let $\{\bar{\xi}_t = (\bar{X}_t^\intercal, \bar{\varepsilon}_t)^\intercal$, $t = 1, ..., n\}$ be an IID sequence that shares the same marginal distribution as ξ_t. For any $t \ne s \ne r$,

$E\left[\phi\left(\bar{\xi}_t,\bar{\xi}_s,\bar{\xi}_r;\gamma\right)\right]=0$. Let $\phi_1(\cdot)=E[\phi\left(\cdot,\bar{\xi}_s,\bar{\xi}_r;\gamma\right)]$ and $\phi_2(\cdot,\cdot;\gamma)=E\left[\phi\left(\cdot,\cdot,\bar{\xi}_r;\gamma\right)\right]$ where $s\neq r$. Let $\bar{\phi}_2(a_1,a_2;\gamma)=\phi_2(a_1,a_2;\gamma)-\phi_1(a_1;\gamma)-\phi_1(a_2;\gamma)$ and $\bar{\phi}_3(a_1,a_2,a_3;\gamma)=\phi(a_1,a_2,a_3;\gamma)-\bar{\phi}_2(a_1,a_2;\gamma)-\bar{\phi}_2(a_1,a_3;\gamma)-\bar{\phi}_2(a_2,a_3;\gamma)$, where a_1, a_2, and a_3 are $(d+1)\times 1$ vectors. By Hoeffding decomposition (e.g., Lee, 1990, p.26), we can $\bar{v}_{nb,1}(\gamma)$ as follows:

$$\bar{v}_{nb,1}(\gamma)=3H_{n(1)}(\gamma)+3H_{n(2)}(\gamma)+H_{n(3)}(\gamma),$$

where $H_{n(1)}(\gamma)=\frac{1}{n}\sum_{t=1}^{n}\phi_1(\xi_t;\gamma)$, $H_{n(2)}(\gamma)=\frac{2}{n(n-1)}\sum_{1\leq t<s\leq n}\bar{\phi}_2(\xi_t,\xi_s;\gamma)$, and $H_{n(3)}(\gamma)=\frac{6}{n(n-1)(n-2)}\sum_{1\leq t<s<r\leq n}\bar{\phi}_3(\xi_t,\xi_s,\xi_r)$. Noting that $\bar{\phi}_2(\cdot,\cdot;\gamma)$ and $\bar{\phi}_3(\cdot,\cdot,\cdot;\gamma)$ are symmetric in its arguments and $E\left[\bar{\phi}_2(a_1,\xi_t;\gamma)\right]=E[\bar{\phi}_3(a_1,\,a_2,\xi_t;\gamma)]=0$, we can readily show that $E\{[H_{n(3)}(\gamma)]^2\}=O\left(n^{-3}b^{-(2d+1)}\right)$, and $E\{[H_{n(2)}(\gamma)]^2\}=O\left(n^{-2}b^{-(d+1)}\right)$, implying that $H_{n(3)}(\gamma)=O_P\left(n^{-3/2}b^{-(2d+1)/2}\right)$ and $H_{n(2)}(\gamma)=O_P\left(n^{-1}b^{-(d+1)/2}\right)$. By modifying the proof of (A.10) in Gozalo and Linton (2001), we can obtain the uniform bounds: $\sup_{\gamma\in\Gamma}\left|H_{n(3)}(\gamma)\right|=O_P\left(n^{-3/2}b^{-(2d+1)/2}\log n\right)=o_P((nb)^{-1/2})$ and $\sup_{\gamma\in\Gamma}\left|H_{n(2)}(\gamma)\right|=O_P\left(n^{-1}b^{-(d+1)/2}\log n\right)=o_P((nb)^{-1/2})$. In addition,

$$\begin{aligned}\phi_1(\xi;\gamma) &= E_{\bar{\xi}_s}E_{\bar{\xi}_r}\left[\phi\left(\xi,\bar{\xi}_s,\bar{\xi}_r;\gamma\right)\right]\\ &= \frac{1}{3b}\{E\left[K_b\left(\bar{X}_s-\bar{X}_r\right)K_b\left(x-\bar{X}_r\right)\left[1\left\{\bar{X}_{s,1}>\gamma\right\}-1\left\{\bar{X}_{r,1}>\gamma\right\}\right]\right]\\ &\quad -E\left[K_b\left(\bar{X}_s-x\right)K_b\left(\bar{X}_r-x\right)\left[1\left\{\bar{X}_{s,1}>\gamma\right\}-1\left\{x_1>\gamma\right\}\right]\right]\}\varepsilon\\ &= \frac{1}{3}\left[c_{1b}(x,\gamma)-c_{2b}(x,\gamma)\right]\varepsilon\end{aligned}$$

where $E_{\bar{\xi}_s}$ denotes expectation with respect to $\bar{\xi}_s$. It follows that

$$\bar{v}_{nb,1}(\gamma)=\frac{1}{n}\sum_{t=1}^{n}\left[c_{1b}(X_t,\gamma)-c_{2b}(X_t,\gamma)\right]\varepsilon_t+o_P((nb)^{-1/2})\text{ uniformly in }\gamma\in\Gamma. \tag{A.9}$$

By (A.5)-(A.9) and the fact that $S_b^{-1}(\gamma)=O(1)$, we have

$$v_{nb}(\gamma)=S_b^{-1}(\gamma)\frac{1}{n}\sum_{t=1}^{n}\left[c_{1b}(X_t,\gamma)-c_{2b}(X_t,\gamma)\right]\varepsilon_t+o_P((nb)^{-1/2})\text{ uniformly in }\gamma\in\Gamma. \tag{A.10}$$

Similarly, we can show that uniformly in $\gamma\in\Gamma$, $\tilde{b}_{nb}(\gamma)\equiv S_{nb}(\gamma)\,b_{nb}(\gamma)=O_P(b^{\upsilon})$ and

$$b_{nb}(\gamma)=O_P(b^{\upsilon}). \tag{A.11}$$

Now, let $\tilde{r}_{nb}(\gamma)\equiv S_{nb}(\gamma)\,r_{nb}(\gamma)$. Let $\psi_b^0(X_t,X_s,X_r;\gamma)=\frac{1}{b}K_b(X_s-X_t)\left[D_s(\gamma)-D_t(\gamma)\right]K_b(X_r-X_t)\left[D_r(\gamma_0)-D_t(\gamma_0)-D_r(\gamma)+D_t(\gamma)\right]$. Let ψ_b denote the symmetric version of ψ_b^0. Then we can write

$$\begin{aligned}\tilde{r}_{nb}(\gamma) &= \beta_0\frac{1}{n^3}\sum_{t=1}^{n}\sum_{s\neq t}^{n}\sum_{r\neq t}^{n}\psi_b^0(X_t,X_s,X_r;\gamma)\\ &= (1+o(1))\frac{6\beta_0}{n(n-1)(n-2)}\sum_{1\leq t<s<r\leq n}\psi_b(X_t,X_s,X_r;\gamma)+O_P(n^{-1}b^{-d})\end{aligned}$$

where the $O_P(n^{-1}b^{-d})$ arises from the $s=r$ terms in the summation. Following the analysis of $\tilde{v}_{nb,1}(\gamma)$ and using Hoeffding decomposition, we can show that

$$\frac{6}{n(n-1)(n-2)}\sum_{1\leq t<s<r\leq n}\psi_b(X_t,X_s,X_r;\gamma)=\theta_{\psi_1}(\gamma)+\frac{3}{n}\sum_{t=1}^{n}\left[\psi_{1b}(X_t;\gamma)-\theta_{\psi_1}(\gamma)\right]+o_P(n^{-1/2}b^{-1/2})$$

where

$$\begin{aligned}
\psi_{1b}(x;\gamma) &= E_{X_1}E_{X_2}[\psi_b(x,X_1,X_2;\gamma)] \\
&= \frac{1}{3b}E_{X_1}E_{X_2}[\psi_b^0(x,X_1,X_2;\gamma)+\psi_b^0(X_1,x,X_2;\gamma)+\psi_b^0(X_2,X_1,x;\gamma)] \\
&= \frac{b}{3}d_b(x,\gamma)[d_b(x,\gamma_0)-d_b(x,\gamma)] \\
&\quad +\frac{1}{3}E\{K_b(X_s-x)[1\{x_1>\gamma\}-1\{X_{s1}>\gamma\}]\cdot[d_b(X_s,\gamma_0)-d_b(X_s,\gamma)]\} \\
&\quad +\frac{1}{3}E\{d_b(X_r,\gamma)\ \ K_b(X_r-x)[1\{x_1>\gamma_0\}-1\{X_{r1}>\gamma_0\}-1\{x_1>\gamma\}+1\{X_{r1}>\gamma\}]\}\} \\
&= \frac{b}{3}\{d_b(x,\gamma)[d_b(x,\gamma_0)-d_b(x,\gamma)]+\bar{d}_b(x;\gamma_0,\gamma)+\bar{d}_b(x;\gamma,\gamma_0)-2\bar{d}_b(x;\gamma,\gamma)\}, \qquad \text{(A.12)}
\end{aligned}$$

$$\theta_{\psi_1}(\gamma)=E[\psi_{1b}(X_t;\gamma)]=bE\{d_b(X_t,\gamma)[d_b(X_t,\gamma_0)-d_b(X_t,\gamma)]\}, \qquad \text{(A.13)}$$

and

$$\bar{d}_b(x;\gamma,\gamma')=\frac{1}{b}E\{d_b(X_t,\gamma)\ \ K_b(X_t-x)[1\{x_1>\gamma'\}-1\{X_{t1}>\gamma'\}]\}.$$

It follows that

$$\tilde{r}_{nb}(\gamma)=\beta_0\left\{\theta_{\psi_1}(\gamma)+\frac{3}{n}\sum_{t=1}^{n}[\psi_{1b}(X_t;\gamma)-\theta_{\psi_1}(\gamma)]\right\}+o_P(n^{-1/2}b^{-1/2})$$

and

$$r_{nb}(\gamma)=\beta_0 S_b^{-1}(\gamma)\left\{\theta_{\psi_1}(\gamma)+\frac{3}{n}\sum_{t=1}^{n}[\psi_{1b}(X_t;\gamma)-\theta_{\psi_1}(\gamma)]\right\}+o_P(n^{-1/2}b^{-1/2}). \qquad \text{(A.14)}$$

Putting (A.4), (A.10), (A.11), and (A.14) together and noticing that $c_{d_b}(\gamma)=S_b^{-1}(\gamma)\,\theta_{\psi_1}(\gamma)$, we have that

$$\begin{aligned}
\hat{\beta}_b(\gamma) &= \beta_0+\beta_0 c_{d_b}(\gamma)+S_b^{-1}(\gamma)\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t,\gamma)-c_{2b}(X_t,\gamma)]\varepsilon_t \\
&\quad +\beta_0 S_b(\gamma)^{-1}\frac{3}{n}\sum_{t=1}^{n}[\psi_{1b}(X_t;\gamma)-\theta_{\psi_1}(\gamma)]+O_P(b^{v})+o_P(n^{-1/2}b^{-1/2})
\end{aligned}$$

uniformly in $\gamma\in\Gamma$ and thus (i) follows.

(ii) This follows from (i) and the fact that $\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t,\gamma)-c_{2b}(X_t,\gamma)]\varepsilon_t$ and $\frac{1}{n}\sum_{t=1}^{n}[\psi_{1b}(X_t;\gamma)$ $-\theta_{\psi_1}(\gamma)]$ are both $O_P\left((nb/\log n)^{-1/2}\right)$ uniformly in $\gamma\in\Gamma$. ■

Lemma A.3 *Suppose that Assumptions A1, A2(i), A3, and A4 hold. Then*

(i) $\sup_{x\in\mathcal{X}_0}\sup_{\gamma\in\Gamma}\|\hat{\alpha}_b(x,\gamma)-\alpha_{0,b}(x,\gamma)\|=O_P(b^{v}+\nu_n)$,

(ii) $\hat{\alpha}_b(x,\gamma_0)-\alpha_0(x)=f(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)\varepsilon_t-f(x)^{-1}c_{0b}(x,\gamma_0)S_b^{-1}(\gamma_0)\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t,\gamma_0)$ $-c_{2b}(X_t,\gamma_0)]\varepsilon_t+O_P(b^{v})+o_P\left((nb)^{-1/2}\right)$ *uniformly in* $x\in\mathcal{X}_0$,

where $\alpha_{0,b}(x,\gamma)=\alpha_0(x)+\delta_{\alpha,b}(x,\gamma)$, *and* $\delta_{\alpha,b}(x,\gamma)=\beta_0 f(x)^{-1}E\{K_b(X_t-x)\ [D_t(\gamma_0)-D_t(\gamma)]\}-$ $\beta_0 c_{d_b}(\gamma)f(x)^{-1}E[K_b(X_t-x)D_t(\gamma)]$.

Proof. (i) Recall that $\hat{\alpha}_b(x;\gamma)=\hat{\alpha}_b(x;\hat{\beta}_b(\gamma),\gamma)$. By (A.1), $\hat{\alpha}_b(x,\gamma)=\alpha_0(x)+V_{nb}(x)+B_{nb}(x)+$ $R_{nb}(x;\hat{\beta}_b(\gamma),\gamma)$, where

$$\begin{aligned}
R_{nb}(x;\hat{\beta}_b(\gamma),\gamma) &= \beta_0\hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)[D_t(\gamma_0)-D_t(\gamma)] \\
&\quad +\left[\beta_0-\hat{\beta}_b(\gamma)\right]\hat{f}_b(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)D_t(\gamma).
\end{aligned}$$

Standard arguments show that uniformly in $x \in \mathcal{X}_0$ and $\gamma \in \Gamma$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x)[D_t(\gamma_0) - D_t(\gamma)] &= E\{K_b(X_t - x)[D_t(\gamma_0) - D_t(\gamma)]\} + O_P(\nu_n), \\ \frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x) D_t(\gamma) &= E\{K_b(X_t - x) D_t(\gamma)\} + O_P(\nu_n),\end{aligned}$$

which, in conjunction with (A.2) and Lemma A.2, implies that

$$\begin{aligned}R_{nb}(x; \hat{\beta}_b(\gamma), \gamma) &= \beta_0 f(x)^{-1} E\{K_b(X_t - x)[D_t(\gamma_0) - D_t(\gamma)]\} \\ &\quad -\beta_0 c_{d_b}(\gamma) f(x)^{-1} E[K_b(X_t - x) D_t(\gamma)] + O_p(b^{\upsilon} + \nu_n) \\ &= \delta_{\alpha,b}(x, \gamma) + O_p(b^{\upsilon} + \nu_n).\end{aligned}$$

Then (i) follows from the standard uniform bounds on $V_{nb}(x)$ and $B_{nb}(x)$ in (A.2).

(ii) By (A.1) and the proofs of Lemmas A.1-A.2, $\hat{\alpha}_b(x; \gamma_0) = \alpha_0(x) + V_{nb}(x) + B_{nb}(x) + R_{nb}(x; \hat{\beta}_b(\gamma_0), \gamma_0)$, where

$$\begin{aligned}R_{nb}\left(x; \hat{\beta}_b(\gamma_0), \gamma_0\right) &= -[\hat{\beta}_b(\gamma_0) - \beta_0] \hat{f}_b(x)^{-1} \frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x) D_t(\gamma_0) \\ &= \left\{-S_b^{-1}(\gamma_0)\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t, \gamma_0) - c_{2b}(X_t, \gamma_0)]\varepsilon_t + O_p(b^{\upsilon}) + o_p\left((nb)^{-1/2}\right)\right\} \\ &\quad \times\left\{f(x)^{-1} c_{0b}(x, \gamma_0) + O_p(b^{\upsilon} + \nu_n)\right\} \\ &= -f(x)^{-1} c_{0b}(x, \gamma_0) S_b^{-1}(\gamma_0) \frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t, \gamma_0) - c_{2b}(X_t, \gamma_0)]\varepsilon_t + o_p\left((nb)^{-1/2}\right).\end{aligned}$$

In addition, by (A.2), the fact that $\frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x)\varepsilon_t = O_P(\nu_n)$ uniformly in $x \in \mathcal{X}_0$, and Assumption A4,

$$V_{nb}(x) = \left[f(x)^{-1} + O_P(b^{\upsilon} + \nu_n)\right]\frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x)\varepsilon_t = f(x)^{-1}\frac{1}{n}\sum_{t=1}^{n} K_b(X_t - x)\varepsilon_t + o_P\left((nb)^{-1/2}\right).$$

Then (ii) follows from the fact that $B_{nb}(x) = O_P(b^{\upsilon})$. ■

To prove the next lemma, we introduce more notations. Let $w_t = w(X_t)$. Recall that $M_n(\gamma, h_b) \equiv \frac{1}{n}\sum_{t=1}^{n}[Y_t - \alpha_b(X_t, \gamma) - \beta_b(\gamma) D_t(\gamma)] w_t$ and $M(\gamma, h_b) \equiv E\{[Y_t - \alpha_b(X_t, \gamma) - \beta_b(\gamma) D_t(\gamma)] w_t\}$. Apparently, $M(\gamma_0, h_{0,b}) = E(\varepsilon_t w_t) = 0$ by the law of iterated expectations. Define

$$\begin{aligned}\mathcal{F}_{1b} &= \{\alpha_b : \alpha_b(\cdot, \gamma) \in C_c^{\lambda}(\mathcal{X}_0) \ \forall \gamma \in \Gamma, \ E\sup_{\gamma' : |\gamma' - \gamma|} |\alpha_b(X_t, \gamma') - \alpha_b(X_t, \gamma)| w_t \leq \bar{c}_{\alpha}|\gamma' - \gamma|\}, \\ \mathcal{F}_{2b} &= \{\beta_b : |\beta_b(\gamma') - \beta_b(\gamma)| \leq \bar{c}_{\beta}|\gamma' - \gamma|\},\end{aligned}$$

where $\bar{c}_{\alpha}$ and $\bar{c}_{\beta}$ are positive constants. Let

$$\mathcal{H} = \mathcal{H}_b = \{(\alpha_b, \beta_b) : \alpha_b \in \mathcal{F}_{1b}, \ \beta_b \in \mathcal{F}_{2b}\}, \tag{A.15}$$

where we suppress the dependence of $\mathcal{H}_b$ on b.

Lemma A.4 *Suppose that Assumptions A1-A4 hold. Let $\delta_n = o(1)$ be an arbitrary positive sequence. Then* $\sup_{\gamma \in \Gamma, \|h_b - h_{0,b}\|_{\mathcal{H}} \leq \delta_n} |M_n(\gamma, h_b) - M(\gamma, h_b)| = o_P(b)$.

Proof. Let $m(X_t;\gamma,h_b) \equiv [\alpha_b(X_t,\gamma)+\beta_b(\gamma)D_t(\gamma)]w_t$. Then

$$M_n(\gamma,h_b)-M(\gamma,h_b)=\frac{1}{n}\sum_{t=1}^{n}[Y_tw_t-E(Y_tw_t)]-\frac{1}{n}\sum_{t=1}^{n}\{m(X_t;\gamma,h_b)-E[m(X_t;\gamma,h_b)]\}$$

By Davydov's and Chebyshev's inequalities, we can readily show that $\frac{1}{n}\sum_{t=1}^{n}[Y_tw_t-E(Y_tw_t)]=O_P(n^{-1/2})$ under Assumptions A1-A2. It suffices to prove (i) by showing that

$$\sup_{\gamma\in\Gamma,\|h_b-h_{0,b}\|_{\mathcal{H}}\leq\delta_n}\left|\frac{1}{n}\sum_{t=1}^{n}\{m(X_t;\gamma,h_b)-E[m(X_t;\gamma,h_b)]\}\right|=o_P(b). \quad \text{(A.16)}$$

The uniform result in (A.16) holds if we can prove the pointwise convergence and then verify the stochastic equicontinuity (s.e.) conditions. The pointwise convergence follows from the direct application of Davydov inequality for strong mixing processes. For the s.e. conditions, we verify the conditions in Lemma 4.2 of Chen (2007). By the C_r-inequality,

$$\begin{aligned}
&|m(X_t;\gamma,h_b)-m(X_t;\gamma',h_b')|^2\\
=\;&|\alpha_b'(X_t,\gamma')-\alpha_b(X_t,\gamma)+\beta_b'(\gamma')D_t(\gamma')-\beta_b(\gamma)D_t(\gamma)|^2w_t^2\\
\leq\;&16\{|\alpha_b'(X_t,\gamma')-\alpha_b(X_t,\gamma')|^2+|\alpha_b(X_t,\gamma')-\alpha_b(X_t,\gamma)|^2+|[\beta_b'(\gamma')-\beta_b(\gamma')]D_t(\gamma')|^2\\
&+|[\beta_b(\gamma')-\beta_b(\gamma)]D_t(\gamma')|^2+|\beta_b(\gamma)[D_t(\gamma')-D_t(\gamma)]|^2\}w_t^2.
\end{aligned}$$

Apparently, uniformly in (γ',h_b') such that $|\gamma'-\gamma|\leq\delta$ and $\|h_b'-h_b\|_{\mathcal{H}}\leq\delta$, the first and third terms are bounded by $16c_w^2\delta^2$ where $c_w=\sup_{x\in\mathcal{X}_0}w(x)$. For the fifth term, using the fact that $|D_t(\gamma')-D_t(\gamma)|=|1\{q_t\leq\gamma'\}-1\{q_t\leq\gamma\}|\leq 1\{|q_t-\gamma|\leq|\gamma'-\gamma|\}$, we have

$$\begin{aligned}
&E\sup_{(\gamma',h_b'):|\gamma'-\gamma|\leq\delta,\|h_b'-h_b\|_{\mathcal{H}}\leq\delta}|\beta_b(\gamma)[D_t(\gamma')-D_t(\gamma)]|^2w_t^2\\
\leq\;&c_w^2c_\beta^2E\sup_{\gamma':\,|\gamma'-\gamma|\leq\delta}1\{|q_t-\gamma|\leq|\gamma'-\gamma|\}\\
\leq\;&c_w^2c_\beta^2E[1\{|q_t-\gamma|\leq\delta\}]=|\beta_b(\gamma)|^2[F_1(\gamma+\delta)-F_1(\gamma-\delta)]\leq 2c_w^2c_\beta^2c_{f_1}\delta,
\end{aligned}$$

where $c_\beta\equiv\sup_{\gamma\in\Gamma}|\beta_b(\gamma)|$. In addition, for any $\alpha_b\in\mathcal{F}_{1b}$ and $\beta_b\in\mathcal{F}_{2b}$, we have

$$\begin{aligned}
&E\sup_{\gamma':|\gamma'-\gamma|\leq\delta}|\alpha_b(X_t,\gamma')-\alpha_b(X_t,\gamma)|^2w_t^2\\
\leq\;&2c_wc_\alpha E\sup_{\gamma':|\gamma'-\gamma|\leq\delta}|\alpha_b(X_t,\gamma')-\alpha_b(X_t,\gamma)|w(X_t)\leq 2c_wc_\alpha\bar{c}_\alpha\delta
\end{aligned}$$

and

$$E\sup_{\gamma':|\gamma'-\gamma|\leq\delta}|[\beta_b(\gamma')-\beta_b(\gamma)]D_t(\gamma')|^2w_t^2\leq 2c_w^2\bar{c}_\beta^2\delta^2$$

where $c_\alpha\equiv\sup_{x\in\mathcal{X}_0}\sup_{\gamma\in\Gamma}|\alpha_b(x,\gamma)|$. It follows that for sufficiently small $\delta>0$

$$\left\{E\left[\sup_{(\gamma',\theta'):|\gamma'-\gamma|\leq\delta,\|h_b'-h_b\|_{\mathcal{H}}\leq\delta}|m(\xi_t;\gamma,h_b)-m(\xi_t;\gamma',h_b')|^2\right]\right\}^{1/2}\leq c_m\delta^{1/2}$$

where c_m is a finite constant that does not depend on δ. This verifies condition (4.2.1) in Chen (2007). Note that Γ is compact and for each $\gamma\in\Gamma$, $\alpha_b(\cdot,\gamma)\in C_c^\lambda(\mathcal{X})$ for $\lambda>d$. The latter implies that condition (4.2.2) in Chen (2007, Lemma 4.2) is also satisfied; see, e.g., Remark 3(ii) in Chen et al. (2003) for the explanation. Condition (4.2.3) in Chen (2007, Lemma 4.2) is ensured by Assumption A1. Consequently, we have (A.16) and the result in (i) holds. ■

The following theorem will be used in the proof Theorem 3.1.

Theorem A.5 *Let* $\hat{h}_b(\cdot,\gamma) = (\hat{\alpha}_b(\cdot,\gamma), \hat{\beta}_b(\gamma))$ *and* $\hat{h}_b = (\hat{\alpha}_b(\cdot,\hat{\gamma}_b), \hat{\beta}_b(\hat{\gamma}_b))$. *Suppose that* $\gamma_0 \in \Gamma$ *satisfies* $M(\gamma_0, h_{0,b}) = 0$, *and that*

(A.1) $\left|M_n(\hat{\gamma}_b, \hat{h}_b)\right| = \inf_{\gamma\in\Gamma}\left|M_n(\gamma, \hat{h}_b(\cdot,\gamma))\right| + o_P(b)$;

(A.2) for all $\delta > 0$, *there exists* $\epsilon(\delta) > 0$ *such that* $\inf_{|\gamma-\gamma_0|>\delta} \frac{1}{b}|M(\gamma, h_{0,b})| \geq \epsilon(\delta)$;

(A.3) $\sup_{\gamma\in\Gamma}|M(\gamma,h_b) - M(\gamma,h_{0,b})| \leq c_M \|h_b - h_{0,b}\|_{\mathcal{H}}$ *for some fixed finite constant* c_M;

(A.4) $\left\|\hat{h}_b - h_{0,b}\right\|_{\mathcal{H}} = o_P(b)$; *and*

(A.5) $\sup_{\gamma\in\Gamma,\|h_b-h_{0,b}\|_{\mathcal{H}}\leq\delta_n}\|M_n(\gamma,h_b) - M(\gamma,h_b)\| = o_P(b)$ *for any positive sequence* $\delta_n = o(1)$.

Then $\hat{\gamma} - \gamma_0 = o_P(1)$.

The above theorem is similar to Theorem 1 in Chen, Linton, and Van Keilegom (2003, CLV hereafter). But the differences are apparent. First, we allow the population objects h_b and $h_{0,b}$ to depend on the bandwidth parameter b. Second, the identification condition in (A.2) reflects the fact that the function $M(\gamma, h_{0,b})$ is flat in the neighborhood of γ_0 so that γ is "weakly identified". In fact, for any fixed $\delta > 0$, there is no way to ensure that $|M(\gamma, h_{0,b})|$ is bounded away from zero uniformly in $\gamma \in \bar{\Gamma}_\delta \equiv \{\gamma' \in \Gamma : |\gamma' - \gamma_0| > \delta\}$, and the division of $M(\gamma, h_{0,b})$ by b helps to achieve identification. Third, condition (A.4) and (A.5) strengthen conditions (1.4) and (1.5′) in CLV and we now have the requirement on the convergence rate of $\hat{h}_b$.

Proof of Theorem A.5. We prove the theorem by modifying the proof of Theorem 1 in CLV, which is similar to that of Corollary (3.2) in Pakes and Pollard (1989). By condition (A.2), for all $\delta > 0$, we have

$$\Pr(|\hat{\gamma} - \gamma_0| \geq \delta) \leq \Pr\left(\frac{1}{b}M(\hat{\gamma}, h_{0,b}) \geq \epsilon(\delta)\right), \tag{A.17}$$

implying that we can prove the theorem by showing that $M(\hat{\gamma}, h_{0,b}) = o_P(b)$. By the triangle inequality,

$$|M(\hat{\gamma}, h_{0,b})| \leq \left|M(\hat{\gamma}, h_{0,b}) - M(\hat{\gamma}, \hat{h}_b)\right| + \left|M(\hat{\gamma}, \hat{h}_b) - M_n(\hat{\gamma}, \hat{h}_b)\right| + \left|M_n(\hat{\gamma}, \hat{h}_b)\right|. \tag{A.18}$$

By conditions (A.3) and (A.4), $|M(\hat{\gamma}, h_{0,b}) - M(\hat{\gamma}, \hat{h}_b)| = o_P(b)$. By conditions (A.4) and (A.5), $|M(\hat{\gamma}, \hat{h}_b) - M_n(\hat{\gamma}, \hat{h}_b)| = o_P(b)$. We are left to show that $|M_n(\hat{\gamma}, \hat{h}_b)| = o_P(b)$. By condition (A.1),

$$\left|M_n(\hat{\gamma}, \hat{h}_b)\right| \leq \inf_{\gamma\in\Gamma}\left|M_n(\gamma, \hat{h}_b)\right| + o_P(b). \tag{A.19}$$

By the triangle inequality, the fact that $M(\gamma_0, h_{0,b}) = 0$, and using conditions (A.3)-(A.5) again, we have uniformly in $\gamma \in \Gamma$

$$\begin{aligned}\left|M_n(\gamma, \hat{h}_b)\right| &\leq \left|M_n(\gamma, \hat{h}_b) - M(\gamma, \hat{h}_b)\right| + \left|M(\gamma, \hat{h}_b) - M(\gamma, h_{0,b})\right| + |M(\gamma, h_{0,b}) - M(\gamma_0, h_{0,b})| \\ &= o_P(b) + o_P(b) + |M(\gamma, h_{0,b}) - M(\gamma_0, h_{0,b})|.\end{aligned}$$

It follows that

$$\inf_{\gamma\in\Gamma}\left|M_n(\gamma, \hat{h}_b)\right| \leq o_P(b) + \inf_{\gamma\in\Gamma}|M(\gamma, h_{0,b}) - M(\gamma_0, h_{0,b})| = o_P(b)$$

and $|M(\hat{\gamma}, h_{0,b})| = o_P(b)$. This completes the proof of Theorem A.5. ■

Proof of Theorem 3.1. We prove the theorem by verifying the conditions in Theorem A.5. (i) is satisfied by (2.20). Noting that $Y_t = \alpha_0(X_t) + \beta_0 D_t(\gamma_0) + \varepsilon_t$, $\alpha_{0,b}(X_t,\gamma) = \alpha_0(X_t) + \delta_{\alpha,b}(X_t,\gamma)$, and $\beta_{0,b}(\gamma) = \beta_0 + \beta_0 c_{d_b}(\gamma)$, we have

$$\begin{aligned}-M(\gamma, h_{0,b}) &= -E\{[Y_t - \alpha_{0,b}(X_t,\gamma) - \beta_{0,b}(\gamma) D_t(\gamma)] w_t\} \\ &= E[\{\delta_{\alpha,b}(X_t,\gamma) + [\beta_{0,b}(\gamma) D_t(\gamma) - \beta_0 D_t(\gamma_0)]\} w_t].\end{aligned}$$

Noting that $E\left[\delta_{\alpha,b}\left(X_t,\gamma\right)w_t\right] = \beta_0 \int E\left\{K_b\left(X_t-x\right)\left[D_t\left(\gamma_0\right)-D_t\left(\gamma\right)\right]\right\}w\left(x\right)dx - \beta_0 c_{d_b}\left(\gamma\right) \int E[K_b(X_t -x)D_t\left(\gamma\right)]w\left(x\right)dx$ and $E\left\{\left[\beta_{0,b}\left(\gamma\right)D_t\left(\gamma\right)-\beta_0 D_t\left(\gamma_0\right)\right]w_t\right\} = \beta_0 E\left\{\left[D_t\left(\gamma\right)-D_t\left(\gamma_0\right)\right]w_t\right\}+\beta_0 c_{d_b}\left(\gamma\right)E[D_t\left(\gamma\right)w_t]$, we have

$$
\begin{aligned}
-M\left(\gamma,h_{0,b}\right) &= \beta_0 \int \left\{E\left\{K_b\left(X_t-x\right)\left[D_t\left(\gamma_0\right)-D_t\left(\gamma\right)\right]\right\}-\left[1\left\{x_1>\gamma_0\right\}-1\left\{x_1>\gamma\right\}\right]f\left(x\right)\right\}w\left(x\right)dx \\
&\quad -\beta_0 c_{d_b}\left(\gamma\right)\left\{\int E\left[K_b\left(X_t-x\right)D_t\left(\gamma\right)\right]-1\left\{x_1>\gamma\right\}f\left(x\right)\right\}w\left(x\right)dx.
\end{aligned}
$$

The first term is

$$
\begin{aligned}
&\beta_0 \int\left\{\int K\left(u\right)\left[1\left\{x_1+bu_1>\gamma_0\right\}-1\left\{x_1+bu_1>\gamma\right\}\right]f\left(x+bu\right)du-\left[1\left\{x_1>\gamma_0\right\}-1\left\{x_1>\gamma\right\}\right]f\left(x\right)\right\} \\
&\times w\left(x\right)dx \\
=\ &\beta_0 \int\int K\left(u\right)\left[1\left\{x_1+bu_1>\gamma_0\right\}-1\left\{x_1+bu_1>\gamma\right\}-1\left\{x_1>\gamma_0\right\}+1\left\{x_1>\gamma\right\}\right]f\left(x+bu\right)du\ w\left(x\right)dx \\
&+\beta_0 \int\left\{\int K\left(u\right)\left[f\left(x+bu\right)-f\left(x\right)\right]du\left[1\left\{x_1>\gamma_0\right\}-1\left\{x_1>\gamma\right\}\right]\right\}w\left(x\right)dx \\
=\ &\beta_0 b\int\left[d_b\left(x,\gamma_0\right)-d_b\left(x,\gamma\right)\right]w\left(x\right)dx+O\left(b^{\upsilon}\right),
\end{aligned}
$$

and the second term is

$$
\begin{aligned}
&\beta_0 c_{d_b}\left(\gamma\right)\left\{\int E\left[K_b\left(X_t-x\right)D_t\left(\gamma\right)\right]-1\left\{x_1>\gamma\right\}f\left(x\right)\right\}w\left(x\right)dx \\
=\ &\beta_0 c_{d_b}\left(\gamma\right)\int\left[\int K\left(u\right)1\left\{x_1+bu_1>\gamma\right\}f\left(x+bu\right)du-1\left\{x_1>\gamma\right\}f\left(x\right)\right]w\left(x\right)dx \\
=\ &\beta_0 c_{d_b}\left(\gamma\right)\int\left[\int K\left(u\right)\left[1\left\{x_1+bu_1>\gamma\right\}-1\left\{x_1>\gamma\right\}\right]f\left(x+bu\right)du\right]w\left(x\right)dx \\
&+\beta_0 c_{d_b}\left(\gamma\right)\int\left[\int K\left(u\right)\left[f\left(x+bu\right)-f\left(x\right)\right]du1\left\{x_1>\gamma\right\}\right]w\left(x\right)dx \\
=\ &\beta_0 c_{d_b}\left(\gamma\right)b\int d_b\left(x,\gamma\right)w\left(x\right)dx+O\left(b^{\upsilon}\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\frac{-1}{b}M\left(\gamma,h_{0,b}\right) &= \beta_0\int\left[d_b\left(x,\gamma_0\right)-d_b\left(x,\gamma\right)\right]w\left(x\right)dx-\beta_0 c_{d_b}\left(\gamma\right)\int d_b\left(x,\gamma\right)w\left(x\right)dx+O\left(b^{\upsilon-1}\right) \\
&= \beta_0\int\left\{\left[d_b\left(x,\gamma_0\right)-d_b\left(x,\gamma\right)\right]-c_{d_b}\left(\gamma\right)d_b\left(x,\gamma\right)\right\}w\left(x\right)dx+O\left(b^{\upsilon-1}\right) \\
&= \beta_0\left[g_w\left(\gamma_0\right)-g_w\left(\gamma\right)\right]-c_{d_b}\left(\gamma\right)g_w\left(\gamma\right)+O\left(b^{\upsilon-1}\right),
\end{aligned}
$$

where $g_w\left(\gamma\right)\equiv\int d_b\left(x,\gamma\right)w\left(x\right)dx$, and $c_{d_b}\left(\gamma\right)\equiv\left(E\left[d_b^2\left(X_t,\gamma\right)\right]\right)^{-1}E\left\{d_b\left(X_t,\gamma\right)\left[d_b\left(X_t,\gamma_0\right)-d_b\left(X_t,\gamma\right)\right]\right\}$ is the solution to the following minimization problem

$$
\min_c E\left\{\left[d_b\left(X_t,\gamma_0\right)-d_b\left(X_t,\gamma\right)\right]-cd_b\left(X_t,\gamma\right)\right\}^2.
$$

To conclude that for any $\delta > 0$, there exists $\epsilon(\delta)$ such that $\left|\frac{1}{b}M(\gamma, h_{0,b})\right| \geq \epsilon(\delta)$ for any γ such that $|\gamma - \gamma_0| \geq \delta$ and $\gamma \in \mathcal{X}_{0,1}$, we do some tedious calculations. Observe that

$$\begin{aligned}
g_w(\gamma) &= \frac{1}{b}\int E\{K_b(X_t - x)[D_t(\gamma) - 1\{x_1 > \gamma\}]\} w(x)\, dx \\
&= \frac{1}{b}\int\int K(u)[1\{x_1 + bu_1 > \gamma\} - 1\{x_1 > \gamma\}] f(x + bu)\, du w(x)\, dx \\
&= \frac{1}{b}\int\int k(u_1)\left[1\{x_1 \leq \gamma\} - 1\left\{u_1 \leq \frac{\gamma - x_1}{b}\right\}\right] du_1 f(x) w(x)\, dx \\
&\quad + \frac{1}{b}\int\int K(u)\left[1\{x_1 \leq \gamma\} - 1\left\{u_1 \leq \frac{\gamma - x_1}{b}\right\}\right][f(x + bu) - f(x)]\, du w(x)\, dx \\
&\equiv g_{w,1}(\gamma) + g_{w,2}(\gamma), \text{ say.}
\end{aligned}$$

For $g_{w,1}(\gamma)$, we have

$$\begin{aligned}
g_{w,1}(\gamma) &= \frac{1}{b}\int_{\mathcal{X}_{0,1}}\left[1\{x_1 - \gamma \leq 0\} - \bar{k}\left(\frac{\gamma - x_1}{b}\right)\right] e_w(x_1)\, dx_1 \\
&= \int\left[1\{u_1 \leq 0\} - \bar{k}(-u_1)\right] e_w(\gamma + bu_1)\, du_1,
\end{aligned}$$

where $\bar{k}(t) = \int_{-1}^{t} k(s)\, ds$, and $e_w(x_1) = \int f(x) w(x)\, dx_{-1}$ has compact support $\mathcal{X}_{0,1}$ that includes γ as an interior point. Noting that $\bar{k}(s) = 0$ for $s \leq -1$ and $= 1$ for $s \geq 1$, we have

$$1\{u_1 \leq 0\} - \bar{k}(-u_1) = 1\{-1 \leq u_1 \leq 0\} - \bar{k}(-u_1) 1\{-1 \leq u_1 \leq 1\}.$$

It follows that $c_g \equiv \int\left[1\{u_1 \leq 0\} - \bar{k}(-u_1)\right] du_1 = 1 - \int_{-1}^{1}\bar{k}(u_1)\, du_1 = 0$, where we have used the fact that $k(\cdot)$ is an even function that is integrated to 1 on its compact support $[-1, 1]$. Consequently,

$$\begin{aligned}
g_{w,1}(\gamma) &= \int[1\{-1 \leq u_1 \leq 0\} - \bar{k}(-u_1) 1\{-1 \leq u_1 \leq 1\}]\left\{e_w(\gamma) + \sum_{j=1}^{\upsilon} e_w^{(j)}(\gamma) u_1^j \frac{b^j}{j!}\right\} du_1 + o(b^\upsilon) \\
&= \sum_{j=1}^{\upsilon}\frac{b^j}{j!} e_w^{(j)}(\gamma) c_{\bar{k},j} + o(b^\upsilon) = O(b),
\end{aligned}$$

where $e_w^{(j)}(\gamma) = \partial^j e_w(\gamma)/\partial\gamma^j$, and $c_{\bar{k},j} = \int_{-1}^{0} u_1^j du_1 - \int_{-1}^{1}\bar{k}(-u_1) u_1^j du_1$. Under Assumption A3, $c_{\bar{k},1} = -\frac{1}{2} + \int_{-1}^{1}\bar{k}(u_1) u_1 du_1 \neq 0$. Analogously, we can show that $g_{w,2}(\gamma) = O\left(b^2\right)$. It follows that

$$\frac{1}{b} g_w(\gamma) = c_{\bar{k},1} e_w^{(1)}(\gamma) + O(b) \text{ and } g_w(\gamma_0) - g_w(\gamma) = O(b) \text{ uniformly in } \gamma.$$

By the calculations in Appendix B and the fact that $c_{d_b}(\gamma_0) = 0$ and that $c_{d_b}(\gamma)$ is continuously differentiable at $\gamma \neq \gamma_0$, there exists $\gamma_\delta^* \in (\gamma, \gamma_0)$ or (γ_0, γ) with $|\gamma - \gamma_0| \geq \delta$ such that

$$|c_{d_b}(\gamma)| = |c_{d_b}(\gamma) - c_{d_b}(\gamma_0)| = |\dot{c}_{d_b}(\gamma_\delta^*)(\gamma - \gamma_0)| \geq |\dot{c}_{d_b}(\gamma_\delta^*)|\,\delta$$

and $b \cdot |\dot{c}_{d_b}(\gamma_\delta^*)|$ is bounded away from zero. Consequently,

$$\begin{aligned}
\frac{1}{b}|M(\gamma, h_{0,b})| &= |c_{d_b}(\gamma) g_w(\gamma)| + O(b) \geq \delta\,|b \cdot \dot{c}_{d_b}(\gamma_\delta^*)|\,\frac{1}{b}|g_w(\gamma)| + O(b) \\
&\geq \delta\,|b \cdot \dot{c}_{d_b}(\gamma_\delta^*)|\left|c_{\bar{k},1} e_w^{(1)}(\gamma)\right|/2,
\end{aligned}$$

as $b \to 0$ and $e_w^{(1)}(\gamma)$ is bounded away from zero on $\bar{\Gamma}_\delta$. Then condition A2 follows.

Condition A3 is satisfied with $c_M = \bar{c}_w \equiv \sup_{x\in\mathcal{X}_0} w(x)$ because

$$
\begin{aligned}
& |M(\gamma, h_b) - M(\gamma, h_{0,b})| \\
= \; & |E\{\{[\alpha_b(X_t,\gamma) - \alpha_{0,b}(X_t,\gamma)] + [\beta_b(\gamma) - \beta_{0,b}(\gamma)] D_t(\gamma)\} w_t\} \\
\leq \; & \bar{c}_w \left\{ \sup_{x\in\mathcal{X}_0} \|\alpha_b(x,\gamma) - \alpha_{0,b}(x,\gamma)\| + |\beta_b(\gamma) - \beta_{0,b}(\gamma)| \right\} \leq \bar{c}_w \|h_b - h_{0,b}\|_{\mathcal{H}}.
\end{aligned}
$$

Condition A4 is satisfied by Lemma A.1 and the fact that $\nu_n = o(b)$ under Assumption A4. Condition A5 holds by Lemma A.3(i). This completes the proof of the theorem. ■

To prove Theorem 3.2, we define some notation. Let $\Gamma_{\delta_{1n}} \equiv \{\gamma \in \Gamma : |\gamma - \gamma_0| \leq \delta_{1n}\}$ and $\mathcal{H}_{\delta_{2n}} \equiv \{h \in \mathcal{H} : \|h - h_{0,b}\|_{\mathcal{H}} \leq \delta_{2n}\}$. For example, given the results in Theorem 3.1 and Assumption A4 that ensures $\nu_n + b^{\upsilon} = o(n^{-\kappa})$, we can take $\delta_{1n} = n^{-\varkappa}$ and $\delta_{2n} = n^{-\kappa}$ with $\varkappa = 0$. For any $(\gamma, h) \in \Gamma_{\delta_{1n}} \times \mathcal{H}_{\delta_{2n}}$, we define the ordinary left and right derivatives of $M(\gamma, h)$ with respect to γ as $\Upsilon_{1b,-}(\gamma, h)$ and $\Upsilon_{1b,+}(\gamma, h)$, respectively. For any $\gamma \in \Gamma_{\delta_{1n}}$, we say that $M(\gamma, h)$ is pathwise differentiable at $h \in \mathcal{H}_{\delta_{2n}}$ in the direction $[\bar{h} - h]$ if $\{h + \tau(\bar{h} - h) : \tau \in (0,1)\} \subset \mathcal{H}$ and

$$
\lim_{\tau\to 0} \left[M\left(\gamma, h(\cdot,\gamma) + \tau(\bar{h}(\cdot,\gamma) - h(\cdot,\gamma))\right) - M(\gamma, h(\cdot,\gamma)) \right] / \tau
$$

exists; and we denote the above limit by $\Upsilon_{2b}(\gamma, h)[\bar{h} - h]$. Define

$$
\begin{aligned}
\mathcal{H}^*_{\delta_{2n}} = & \left\{ h = (\alpha, \beta) \in \mathcal{H}_{\delta_{2n}} : \sup_{x\in\mathcal{X}_0} \frac{1}{b} |\alpha(x,\gamma) - \alpha_{0,b}(x,\gamma) - \alpha(x,\gamma_0) + \alpha_{0,b}(x,\gamma_0)| = o(|\gamma - \gamma_0|) \right. \\
& \left. \text{and } \frac{1}{b} |\beta(\gamma) - \beta_{0,b}(\gamma) - \beta(\gamma_0) + \beta_{0,b}(\gamma_0)| = o(|\gamma - \gamma_0|) \;\; \forall \, \gamma \in \Gamma \right\}.
\end{aligned}
$$

Below, we further restrict our attention to a subclass of $\mathcal{H}_{\delta_{2n}} : \bar{\mathcal{H}}_{\delta_{2n}} = \mathcal{H}_{\delta_{2n}} \cap \mathcal{H}^*_{\delta_{2n}}$.

Lemma A.6 *Suppose that Assumptions A1-A4 hold. Let $\varkappa \in [0, \frac{1}{2} \wedge (\kappa(1 - \frac{\lambda}{2d}) + \eta))$. Then* $\sup_{|\gamma-\gamma_0|\leq n^{-\varkappa}} \sup_{h_b\in\bar{\mathcal{H}}_{\delta_{2n}}, \|h_b - h_{0,b}\|_\infty \leq n^{-\kappa}} |M_n(\gamma, h_b) - M(\gamma, h_b) - M_n(\gamma_0, h_{0,b})| = o_P\left(n^{-1/2}\vartheta_n \log n\right)$, *where $\vartheta_n = n^{-\varkappa} b^{-1} \wedge 1$, and $a \wedge b = \min(a, b)$.*

Proof. Recall that $m(X_t; \gamma, h_b) = [\alpha_b(X_t,\gamma) + \beta_b(\gamma) D_t(\gamma)] w_t$. Noting that $M(\gamma_0, h_{0,b}) = 0$ and $m(X_t, \gamma_0, h_{0,b}) = [\alpha_{0,b}(X_t,\gamma) + \beta_{0,b}(\gamma) D_t(\gamma)] w_t$, we have

$$
\begin{aligned}
& -[M_n(\gamma, h_b) - M(\gamma, h_b) - M_n(\gamma_0, h_{0,b})] \\
= \; & \frac{1}{n} \sum_{t=1}^{n} \{ m(X_t; \gamma, h_b) - m(X_t; \gamma_0, h_{0,b}) - E[m(X_t; \gamma, h_b) - m(X_t; \gamma_0, h_{0,b})] \} \\
= \; & \frac{1}{n} \sum_{t=1}^{n} \{\Delta_t(\alpha,\gamma) - E[\Delta_t(\alpha,\gamma)]\} + \frac{1}{n} \sum_{t=1}^{n} \{\Lambda_t(\beta,\gamma) - E[\Lambda_t(\beta,\gamma)]\}
\end{aligned}
$$

where $\Delta_t(\alpha,\gamma) = [\alpha_b(X_t,\gamma) - \alpha_{0,b}(X_t,\gamma_0)] w_t$, and $\Lambda_t(\beta,\gamma) = [\beta_b(\gamma) D_t(\gamma) - \beta_{0,b}(\gamma_0) D_t(\gamma_0)] w_t$. It suffices to show that

$$
\sup_{|\gamma-\gamma_0|\leq n^{-\varkappa}, \|\alpha - \alpha_{0,b}\|_\infty \leq n^{-\kappa}} \left| \frac{1}{n} \sum_{t=1}^{n} \{\Delta_t(\alpha,\gamma) - E[\Delta_t(\alpha,\gamma)]\} \right| = o_P\left(n^{-1/2}\vartheta_n \log n\right), \tag{A.20}
$$

and

$$
\sup_{|\gamma-\gamma_0|\leq n^{-\varkappa}, \; |\beta - \beta_{0,b}|_\infty \leq n^{-\kappa}} \left| \frac{1}{n} \sum_{t=1}^{n} \{\Lambda_t(\beta,\gamma) - E[\Lambda_t(\beta,\gamma)]\} \right| = o_P\left(n^{-1/2}\vartheta_n \log n\right). \tag{A.21}
$$

We only prove (A.20) as the proof of (A.21) is simpler.

Let $\Gamma_n = \{\gamma : |\gamma - \gamma_0| \leq n^{-\varkappa}\}$ and $\mathcal{A}_n = \{\alpha : \alpha(\cdot,\gamma) \in C_c^\lambda(\mathcal{X}_0)\ \forall\gamma \in \Gamma_n,\ \|\alpha - \alpha_{0,b}\|_\infty \leq n^{-\kappa}\}$. We first create a grid using regions of the form $\Gamma_{j,n} = \{\gamma : |\gamma - \gamma_j| \leq n^{-1/2} b\vartheta_n\}$. By selecting γ_j to lay on the grid, Γ_n can be covered with $N_1 = \lfloor n^{1/2-\varkappa} b^{-1} \vartheta_n^{-1} \rfloor + 1 = O(n^{\frac{1}{2}+(\eta-\varkappa)_+})$ such regions $\Gamma_{j,n}$ for $j = 1, ..., N_1$, where $a_+ = \max(a, 0)$.

Let $\bar{\Delta}_t(\alpha,\gamma) = \Delta_t(\alpha,\gamma) - E[\Delta_t(\alpha,\gamma)]$. Using $|\alpha_b(x,\gamma) - \alpha_b(x,\gamma_j)| \leq c\,|\gamma - \gamma_j|/b$, we have

$$\begin{aligned}
\sup_{\gamma\in\Gamma_j}\left|\frac{1}{n}\sum_{t=1}^n \{\Delta_t(\alpha,\gamma) - E[\Delta_t(\alpha,\gamma)]\}\right| &\leq \left|\frac{1}{n}\sum_{t=1}^n \{\bar{\Delta}_t(\alpha,\gamma_j)\}\right| + \sup_{\gamma\in\Gamma_j}\left|\frac{1}{n}\sum_{t=1}^n \{\bar{\Delta}_t(\alpha,\gamma) - \bar{\Delta}_t(\alpha,\gamma_j)\}\right| \\
&\leq \left|\frac{1}{n}\sum_{t=1}^n \bar{\Delta}_t(\alpha,\gamma_j)\right| + O_P\left(n^{-1/2}\vartheta_n\right)
\end{aligned}$$

uniformly in j and α. Then we can prove (A.20) by showing that

$$\sup_{\|\alpha-\alpha_{0,b}\|_\infty \leq n^{-\kappa}} \max_{1\leq j\leq N_1} \left|\frac{1}{n}\sum_{t=1}^n \bar{\Delta}_t(\alpha,\gamma_j)\right| = o_P(n^{-1/2}\vartheta_n \log n). \tag{A.22}$$

To show (A.22), we follow the proof of Lemma 1 in Mammen et al. (2012) and apply a chaining argument. For $s \geq 0$, let $\mathcal{A}_{s,n}^*$ be a set of functions chosen such that for each $\alpha \in \mathcal{A}_n$, there exists $\alpha_s \in \mathcal{A}_{s,n}^*$ such that $\|\alpha - \alpha_s\|_\infty \leq 2^{-s} n^{-\kappa}$. That is, the functions in $\mathcal{A}_{s,n}^*$ are the midpoints of a $(2^{-s}n^{-\kappa})$-covering of $\mathcal{A}_n$. Under our conditions, the set $\mathcal{A}_{s,n}^*$ can be chosen such that its cardinality $\#\mathcal{A}_{s,n}^*$ is at most $\lfloor C\exp((2^{-s}n^{-\kappa})^{-\varsigma_1} n^{\varsigma_2})\rfloor$ for some $C > 0$, where $\varsigma_1 = d/\lambda$ and $\varsigma_2 > 0$ is an arbitrarily small number (see the discussion after Assumption 3 in Mammen et al. (2012)). For any $\alpha \in \mathcal{A}_n$, we now choose $\alpha_s \in \mathcal{A}_{s,n}^*$ such that $\|\alpha_s - \alpha\|_\infty \leq 2^{-s} n^{-\kappa}$ for $s = 0, 1, ..., N_2 = O(\log n)$. We consider the chain

$$\bar{\Delta}_t(\alpha,\gamma_j) = \bar{\Delta}_t(\alpha_0,\gamma_j) - \sum_{s=1}^{N_2}\left[\bar{\Delta}_t(\alpha_{s-1},\gamma_j) - \bar{\Delta}_t(\alpha_s,\gamma_j)\right] - \bar{\Delta}_t(\alpha_{N_2},\gamma_j).$$

It suffices to prove (A.22) by showing that

$$P\left(\max_{1\leq j\leq N_1}\left|\frac{1}{n}\sum_{t=1}^n \bar{\Delta}_t(\alpha_0,\gamma_j)\right| \geq n^{-1/2}\vartheta_n \log n\right) = o(1), \tag{A.23}$$

$$P\left(\sup_{\|\alpha-\alpha_{0,b}\|_\infty \leq n^{-\kappa}} \max_{1\leq j\leq N_1}\left|\frac{1}{n}\sum_{t=1}^n\sum_{s=1}^{N_2}\left[\bar{\Delta}_t(\alpha_{s-1},\gamma_j) - \bar{\Delta}_t(\alpha_s,\gamma_j)\right]\right| \geq n^{-1/2}\vartheta_n \log n\right) = o(1), \tag{A.24}$$

and

$$P\left(\max_{1\leq j\leq N_1}\left|\frac{1}{n}\sum_{t=1}^n \bar{\Delta}_t(\alpha_{N_2},\gamma_j)\right| \geq n^{-1/2}\vartheta_n \log n\right) = o(1). \tag{A.25}$$

For (A.23), noting that $\left|\bar{\Delta}_t(\alpha_0,\gamma_j)\right| \leq c_\Delta \vartheta_n$ for some finite $c_\Delta > 0$ for each (α_0,γ_j) such that $|\gamma_j - \gamma_0| \leq n^{-\varkappa}$ and $\|\alpha_0 - \alpha_{0,b}\|_\infty \leq n^{-\kappa}$, we can apply Billingsley's inequality (e.g., Bosq (1998, p.20) to show that

$$\begin{aligned}
\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n \bar{\Delta}_t(\alpha_0,\gamma_j)\right) &= \frac{1}{n}\sum_{t=1}^n \mathrm{Var}\left(\bar{\Delta}_t(\alpha_0,\gamma_j)\right) + \frac{2}{n}\sum_{t=1}^{n-1}\sum_{s=t+1}^n \mathrm{Cov}\left(\bar{\Delta}_t(\alpha_0,\gamma_j), \bar{\Delta}_s(\alpha_0,\gamma_j)\right) \\
&\leq c_\Delta^2 \vartheta_n^2 \left\{1 + \frac{8}{n}\sum_{t=1}^{n-1}\sum_{s=t+1}^n \beta_{s-t}\right\} \leq c_{\bar{\Delta}} \vartheta_n^2,
\end{aligned}$$

where $c_{\bar{\Delta}} = c_{\Delta}^2(1+8\sum_{s=1}^{\infty}\beta_s) < \infty$. By Bernstein's inequality for strong mixing processes with geometric decay rate (e.g., Merlevede et al. (2009, Theorem 2)),

$$\begin{aligned}
& P\left(\max_{1\le j\le N_1}\left|\frac{1}{n}\sum_{t=1}^{n}\bar{\Delta}_t\left(\alpha_0,\gamma_j\right)\right| \ge n^{-1/2}\vartheta_n\log n\right) \\
\le\ & \sum_{j=1}^{N_1} P\left(\left|\sum_{t=1}^{n}\bar{\Delta}_t\left(\alpha_0,\gamma_j\right)\right| \ge n^{1/2}\vartheta_n\log n\right) \\
\le\ & N_1\exp\left(-\frac{Cn\vartheta_n^2\left(\log n\right)^2}{nc_{\bar{\Delta}}\vartheta_n^2 + c_{\Delta}^2\vartheta_n^2 + \left(n^{1/2}\vartheta_n\log n\right)c_{\Delta}\vartheta_n\left(\log n\right)^2}\right) \to 0 \text{ as } n\to\infty.
\end{aligned}$$

Analogously, we can prove (A.25).

To prove (A.24), observe that

$$\begin{aligned}
& \Pr\left(\sup_{\left\|\alpha-\alpha_{0,b}\right\|_{\infty}\le n^{-\kappa}}\max_{1\le j\le N_1}\left|\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{N_2}\left[\bar{\Delta}_t\left(\alpha_{s-1},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s,\gamma_j\right)\right]\right| \ge n^{-1/2}\vartheta_n\log n\right) \\
\le\ & \sum_{s=1}^{N_2}\sum_{j=1}^{N_1}\Pr\left(\sup_{\alpha\in\mathcal{A}_n}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s_1-1},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_{s_1},\gamma_j\right)\right]\right| \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
\le\ & \sum_{s=1}^{N_2}\sum_{j=1}^{N_1}\#\mathcal{A}_{s-1,n}^{*}\#\mathcal{A}_{s,n}^{*}\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1}^{*},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s^{**},\gamma_j\right)\right] \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
& +\sum_{s=1}^{N_2}\sum_{j=1}^{N_1}\#\mathcal{A}_{s-1,n}^{*}\#\mathcal{A}_{s,n}^{*}\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\tilde{\alpha}_s^{*},\gamma_j\right)-\bar{\Delta}_t\left(\tilde{\alpha}_s^{**},\gamma_j\right)\right] < c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
\equiv\ & T_1+T_2,\ \text{say},
\end{aligned}$$

where $c_a = \sum_{s=1}^{N_2}2^{-as}$, $a>0$, and $\alpha_{s-1}^{*},\tilde{\alpha}_{s-1}^{*}\in\mathcal{A}_{s-1,n}^{*}$ and $\alpha_s^{**},\tilde{\alpha}_s^{**}\in\mathcal{A}_{s,n}^{*}$ are chosen such that

$$\begin{aligned}
& \Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1}^{*},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s^{**},\gamma_j\right)\right] \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
=\ & \max_{\alpha_{s-1},\alpha_s}\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s,\gamma_j\right)\right] \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right)
\end{aligned}$$

and

$$\begin{aligned}
& \Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\tilde{\alpha}_s^{*},\gamma_j\right)-\bar{\Delta}_t\left(\tilde{\alpha}_s^{**},\gamma_j\right)\right] < c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
=\ & \max_{\alpha_{s-1},\alpha_s}\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s,\gamma_j\right)\right] \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right).
\end{aligned}$$

We now show that $T_1 = o(1)$.

$$\begin{aligned}
T_1 =\ & \sum_{s=1}^{N_2}\sum_{j=1}^{N_1}\#\mathcal{A}_{s-1,n}^{*}\#\mathcal{A}_{s,n}^{*}\Pr\left(\frac{1}{n}\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1}^{*},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s^{**},\gamma_j\right)\right] \ge c_a2^{-as}n^{-1/2}\vartheta_n\log n\right) \\
\le\ & C\sum_{s=1}^{N_2}\sum_{j=1}^{N_1}\exp\left\{\left(1+2^{-\varsigma_1}\right)\left(2^{-s}n^{-\kappa}\right)^{-\varsigma_1}n^{\varsigma_2}\right\} \\
& \times\Pr\left(\left|\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1}^{*},\gamma_j\right)-\bar{\Delta}_t\left(\alpha_s^{**},\gamma_j\right)\right]\right| \ge c_a2^{-as}n^{1/2}\vartheta_n\log n\right).
\end{aligned}$$

Noting that $\left|\bar{\Delta}_t\left(\alpha_{s-1}^*, \gamma_j\right) - \bar{\Delta}_t\left(\alpha_s^{**}, \gamma_j\right)\right| \leq c_\Delta 2^{-s} n^{-\kappa}$, we have by Bernstein's inequality,

$$
\begin{aligned}
& \Pr\left(\left|\sum_{t=1}^{n}\left[\bar{\Delta}_t\left(\alpha_{s-1}^*, \gamma_j\right) - \bar{\Delta}_t\left(\alpha_s^{**}, \gamma_j\right)\right]\right| \geq c_a 2^{-as} n^{1/2} \vartheta_n \log n\right) \\
\leq \; & \exp\left(-\frac{C c_a^2 2^{-2as} n \vartheta_n^2 (\log n)^2}{n c_{\bar{\Delta}} 2^{-2s} n^{-2\kappa} + c_\Delta^2 2^{-2s} n^{-2\kappa} + c_a 2^{-as} n^{1/2} \vartheta_n \log n \left(c_\Delta 2^{-s} n^{-\kappa}\right) (\log n)^2}\right) \\
\leq \; & \exp\left(-C c_a^2 c_{\bar{\Delta}}^{-1} 2^{2s(1-a)} n^{2\kappa} \vartheta_n^2 (\log n)^2 / 2\right) \text{ for sufficiently large } n.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
T_1 \;\leq\; & C N_1 \sum_{s=1}^{N_2} \exp\left(2^{1+s\varsigma_1} n^{\kappa\varsigma_1+\varsigma_2} - C c_a^2 c_{\bar{\Delta}}^{-1} 2^{2s(1-a)} n^{2\kappa} \vartheta_n^2 (\log n)^2 / 2\right) \\
\leq\; & C \sum_{s=1}^{N_2} \exp\left(2^{1+s\varsigma_1} n^{\kappa\varsigma_1+\varsigma_2} - 2c^* 2^{2s(1-a)} n^{2\kappa} \vartheta_n^2 (\log n)^2 + \left(\frac{1}{2} + (\eta - \varkappa)_+\right) \log n\right) \\
\leq\; & C \sum_{s=1}^{N_2} \exp\left(-c^* 2^{2s(1-a)} n^{2\kappa} \vartheta_n^2 (\log n)^2\right) \to 0 \text{ as } n \to \infty,
\end{aligned}
$$

where $c^* = C c_a^2 c_{\bar{\Delta}}^{-1}/4$ and the third inequality follows from the fact that $2^{s\varsigma_1} \leq 2^{2s(1-a)}$ when $a > 0$ is small enough, $n^{2\kappa}\vartheta_n^2 \propto n^{2\kappa+2(\eta-\chi)_-}$ with $a_- = \min(a, 0)$, and $\kappa\varsigma_1 < 2\kappa + 2(\eta - \chi)_-$ under our assumption. Consequently, $T_1 = o(1)$. Similarly, $T_2 = o(1)$. We have proved (A.24). ■

To prove Theorem 3.2 on the basis of Theorem 3.1, we first apply Lemma A.6 with $\varkappa = 0$ to prove an intermediate result for $\hat{\gamma} : \hat{\gamma} - \gamma_0 = o_P\left(n^{-1/2} \log n\right)$. Given this new result, we can apply Lemma A.6 with $\varkappa \in \left(\frac{3}{2}\eta, \frac{1}{2} \wedge (\kappa(1 - \frac{\lambda}{2d}) + \eta\right)$ to obtain the desired rate of consistency: $\hat{\gamma} - \gamma_0 = O_P\left((n/b)^{-1/2}\right)$.

The following theorem will be used in the proof Theorem 3.2.

Theorem A.7 *Let $\delta_{1n} = n^{-\varkappa}$ and $\delta_{2n} = o(n^{-\kappa})$. Suppose that $\gamma_0 \in \Gamma_{\delta_{1n}}$ satisfies $M(\gamma_0, h_{0,b}) = 0$ and that $\hat{\gamma} - \gamma_0 = o_P(n^{-\varkappa})$. Suppose that*

(B.1) $\left|M_n(\hat{\gamma}, \hat{h})\right| = \inf_{\gamma \in \Gamma_{\delta_{1n}}} \left|M_n(\gamma, \hat{h})\right| + o_P\left((n/b)^{-1/2}\right)$.

(B.2) (i) Let $\Upsilon_{1b,-}(\gamma, h_{0,b})$ and $\Upsilon_{1b,+}(\gamma, h_{0,b})$ denote the ordinary left and right derivatives of $M(\gamma, h_{0,b})$ with respect to γ, respectively. $\Upsilon_{1b,-}(\gamma, h_{0,b})$ and $\Upsilon_{1b,+}(\gamma, h_{0,b})$ exist for all $\gamma \in \Gamma_{\delta_{1n}}$. (ii) $\Gamma_{1b,-}(\gamma_0, h_{0,b})$ and $\Gamma_{1b,+}(\gamma_0, h_{0,b})$ are continuous at $\gamma = \gamma_0$ and bounded away from zero and infinity as $n \to \infty$.

(B.3) For all $\gamma \in \Gamma_{\delta_{1n}}$, the pathwise derivative $\Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}]$ of $M(\gamma, h_{0,b})$ exists in all directions $[h - h_{0,b}] \in \mathcal{H}$; and for all $(\gamma, h) \in \Gamma_{\tilde{\delta}_{1n}} \times \bar{\mathcal{H}}_{\tilde{\delta}_{2n}}$ with positive sequences $\tilde{\delta}_{1n} = o(1)$ and $\tilde{\delta}_{2n} = o(b)$: (i) $|M(\gamma, h) - M(\gamma, h_{0,b}) - \Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}]| \leq c_{Mb}^2 \|h - h_{0,b}\|_{\mathcal{H}}^2$ for some constant c_{Mb} that may depend on b; (ii) $\|\Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}] - \Upsilon_{2b}(\gamma_0, h_{0,b})[h - h_{0,b}]\| \leq o(1)\, \tilde{\delta}_{1n}$.

(B.4) $\hat{h} \in \mathcal{H}_{\delta_{2n}}$ w.p.a.1, and $c_{Mb}\left\|\hat{h} - h_{0,b}\right\|_{\mathcal{H}} = o_P\left((n/b)^{-1/4}\right)$.

(B.5) $\sup_{|\gamma - \gamma_0| \leq n^{-\varkappa}, \|h - h_{0,b}\|_{\mathcal{H}} \leq n^{-\kappa}} |M_n(\gamma, h) - M(\gamma, h) - M_n(\gamma_0, h_{0,b})| = o_P\left(n^{-1/2}\vartheta_n \log n\right)$.

(B.6) $\sqrt{n/b}\left\{M_n(\gamma_0, h_{0,b}) + \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}]\right\} \xrightarrow{d} N(0, V_\gamma)$ for some $V_\gamma > 0$.

Then (i) $\hat{\gamma} - \gamma_0 = O_P\left(n^{-1/2} \log n\right)$ if $\varkappa = 0$ in (B.5); (ii) $\sqrt{n/b}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, \Omega_\gamma)$ where $\Omega_\gamma \equiv \lim_{n \to \infty} \Upsilon_{1b}^{-1} V_\gamma \Upsilon_{1b}^{-1}$ if $\varkappa > \frac{3}{2}\eta$ in (B.5).

The above theorem is similar to Theorem 2 in CLV. The major differences lie in three aspects. (i) The population objects such as $\Upsilon_{1b}(\cdot, \cdot)$ and $\Upsilon_{2b}(\gamma, h)[\bar{h} - h]$, are now allowed to depend on the bandwidth parameter b. (ii) The order of the remainder term in the linear expansion of $M(\gamma, h)$ with respect to its second argument may depend on $\|h - h_{0,b}\|_{\mathcal{H}}^2$ and b as well, and as a result, we require

$c_{Mb}^2 \|h - h_{0,b}\|_{\mathcal{H}}^2 = o_P\left((n/b)^{-1/2}\right)$ in condition (B.4). (ii) The stochastic equicontinuity result in condition (B.5) is required to be satisfied only for γ and h that are sufficiently close to the population truth. (iii) Despite the fact that condition (B.6) implies that $M_n(\gamma_0, h_{0,b}) + \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}] = O_P((n/b)^{-1/2})$, it does not imply each term in the last summation is $O_P((n/b)^{-1/2})$, and in fact it is easy to check that $M_n(\gamma_0, h_{0,b}) = O_P(n^{-1/2}b^{-1/2})$ in our case.

Proof of Theorem A.7. We prove the theorem by following the proof of Theorem 2 in CLV closely. By assumption

$$\Pr\left(|\hat{\gamma} - \gamma_0| \geq \delta_{1n}, \left\|\hat{h}_b - h_{0,b}\right\|_{\mathcal{H}} \geq \delta_{2n}\right) \to 0. \tag{A.26}$$

So we only focus on $(\gamma, h) \in \Gamma_{\delta_{1n}} \times \bar{\mathcal{H}}_{\delta_{2n}}$. By condition (B.2) and the fact that $M(\gamma_0, h_{0,b}) = 0$, we have

$$\hat{\gamma} - \gamma_0 = \left\{\left[\Gamma_{1b,+}\left(\gamma_+^*, h_{0,b}\right)\right]^{-1} 1\{\hat{\gamma} > \gamma_0\} + \left[\Gamma_{1b,-}\left(\gamma_-^*, h_{0,b}\right)\right]^{-1} 1\{\hat{\gamma} < \gamma_0\}\right\} M(\hat{\gamma}, h_{0,b}) \tag{A.27}$$

where γ_+^* and γ_-^* lie between $\hat{\gamma}$ and γ_0 and $\left|\Gamma_{1b,\pm}\left(\gamma_\pm^*, h_{0,b}\right)^{-1}\right| \xrightarrow{p} \lim_{n\to\infty}\left|\Gamma_{1b,\pm}(\gamma_0, h_{0,b})^{-1}\right| > 0$. For notational convenience, we frequently write $\hat{\gamma}$ as $\hat{\gamma}_+$ if it is larger than γ_0 and $\hat{\gamma}_-$ otherwise; we use $\hat{\gamma}_\pm$ (c.f., $\gamma_\pm^*$) to denote either $\hat{\gamma}_+$ or $\hat{\gamma}_-$ (c.f., γ_+^* or γ_-^*), which will be clear from the context. Then (A.27) can be rewritten as $\hat{\gamma}_\pm - \gamma_0 = \left[\Gamma_{1b,\pm}\left(\gamma_\pm^*, h_{0,b}\right)\right]^{-1} M\left(\hat{\gamma}_\pm, h_{0,b}\right)$. The probability order of $\hat{\gamma}_\pm - \gamma_0$ is then determined by that of $\left|M\left(\hat{\gamma}_\pm, h_{0,b}\right)\right|$.

By the triangle inequality,

$$\begin{aligned}
\left|M\left(\hat{\gamma}_\pm, h_{0,b}\right)\right| \leq\ & \left|M(\hat{\gamma}_\pm, h_{0,b}) - M(\hat{\gamma}_\pm, \hat{h}_b) + \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}]\right| \\
& + \left|M(\hat{\gamma}_\pm, \hat{h}_b) - M_n(\hat{\gamma}_\pm, \hat{h}_b) + M_n(\gamma_0, h_{0,b})\right| \\
& + \left|M_n(\hat{\gamma}_\pm, \hat{h}_b)\right| + \left|M_n(\gamma_0, h_{0,b}) + \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}]\right|.
\end{aligned} \tag{A.28}$$

The first term on the rhs of (A.28) is bounded from above by

$$\begin{aligned}
& \left|\Upsilon_{2b}\left(\hat{\gamma}_\pm, h_{0,b}\right)[\hat{h}_b - h_{0,b}] - \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}]\right| \\
& + \left|M(\hat{\gamma}_\pm, \hat{h}_b) - M(\hat{\gamma}, h_{0,b}) - \Upsilon_{2b}\left(\hat{\gamma}_\pm, h_{0,b}\right)[\hat{h}_b - h_{0,b}]\right| \\
=\ & \left(\hat{\gamma}_\pm - \gamma_0\right) o_P(1) + O_P\left(c_M^2 \left\|\hat{h}_b - h_{0,b}\right\|_{\mathcal{H}}^2\right) \\
=\ & \left[\Gamma_{1b,\pm}(\gamma_0, h_{0,b})\right]^{-1} M\left(\hat{\gamma}_\pm, h_{0,b}\right) o_P(1) + O_P\left((n/b)^{-1/2}\right)
\end{aligned} \tag{A.29}$$

by conditions (B.3), (B.4), and (B.6), and (A.27). By conditions (B.5) and (B.6) and Theorem 3.1,

$$\left|M(\hat{\gamma}_\pm, \hat{h}_b) - M_n(\hat{\gamma}_\pm, \hat{h}_b) + M_n(\gamma_0, h_{0,b})\right| = o_P\left(n^{-1/2}\vartheta_n \log n\right) \tag{A.30}$$

and

$$M_n(\gamma_0, h_{0,b}) + \Upsilon_{2b}(\gamma_0, h_{0,b})[\hat{h}_b - h_{0,b}] = O_P\left((n/b)^{-1/2}\right). \tag{A.31}$$

Then conditions (B.1)-(B.2) and (A.28)-(A.31) imply that

$$\begin{aligned}
\left|M\left(\hat{\gamma}_\pm, h_{0,b}\right)\right| \times \{1 - o_P(1)\} \leq\ & \left|M_n(\hat{\gamma}_\pm, \hat{h}_b)\right| + O_P\left((n/b)^{-1/2}\right) + o_P\left(n^{-1/2}\vartheta_n \log n\right) \\
\leq\ & \inf_{\gamma \in \Gamma_{\delta_{1n}}} \left|M_n(\gamma, \hat{h}_b)\right| \{1 + o_P(1)\} + O_P((n/b)^{-1/2}) \\
& + o_P\left(n^{-1/2}\vartheta_n \log n\right).
\end{aligned} \tag{A.32}$$

Again, under conditions (B.3)-(B.6), we have that uniformly in γ

$$\begin{aligned}\left|M_n(\gamma,\hat{h}_b)\right| &\leq \left|M_n(\gamma,\hat{h}_b)-M(\gamma,\hat{h}_b)-M_n(\gamma_0,h_{0,b})\right|+\left|M(\gamma,\hat{h}_b)-M(\gamma,h_{0,b})-\Upsilon_{2b}\left(\gamma,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right| \\ &\quad +\left|M(\gamma,h_{0,b})\right|+\left|\Upsilon_{2b}\left(\gamma,h_{0,b}\right)[\hat{h}_b-h_{0,b}]-\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right| \\ &\quad +\left|M_n(\gamma_0,h_{0,b})+\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right| \\ &= o_P\left(n^{-1/2}\vartheta_n\log n\right)+o_P\left((n/b)^{-1/2}\right)+\left|M(\gamma,h_{0,b})\right|+\left|\gamma-\gamma_0\right|o_P\left(1\right)+O_P\left((n/b)^{-1/2}\right).\end{aligned}$$

This, in conjunction with (A.32), the fact that $M(\gamma_0,h_{0,b})=0$, and that $\gamma_0\in\Gamma_{\delta_{1n}}$, implies that

$$\begin{aligned}\inf_{\gamma\in\Gamma_{\delta_{1n}}}\left|M_n(\gamma,\hat{h}_b)\right| &\leq \inf_{\gamma\in\Gamma_{\delta_{1n}}}\left\{\left|M(\gamma,h_{0,b})-M(\gamma_0,h_{0,b})\right|+\left|\gamma-\gamma_0\right|o_P\left(1\right)\right\}+O_P\left((n/b)^{-1/2}\right) \\ &= O_P\left((n/b)^{-1/2}\right). \qquad \text{(A.33)}\end{aligned}$$

Then by (A.32) and (A.26), $\left|M\left(\hat{\gamma}_\pm,h_{0,b}\right)\right|=O_P\left((n/b)^{-1/2}\right)+o_P\left(n^{-1/2}\vartheta_n\log n\right)$ and $\hat{\gamma}_\pm-\gamma_0=O\left(\left|M\left(\hat{\gamma}_\pm,h_{0,b}\right)\right|\right)=O_P\left((n/b)^{-1/2}\right)+o_P\left(n^{-1/2}\vartheta_n\log n\right)$. The first part of the theorem follows by noticing that $O_P\left((n/b)^{-1/2}\right)+o_P\left(n^{-1/2}\vartheta_n\log n\right)=o_P\left(n^{-1/2}\log n\right)$ in the case where $\varkappa=0$.

In view of the condition that $b\propto n^{-\eta}$, we have $n^{-1/2}\vartheta_n\log n=o\left((n/b)^{-1/2}\right)$ and $\hat{\gamma}_\pm-\gamma_0=O_P\left((n/b)^{-1/2}\right)$ in the case where $\varkappa>3\eta/2$. To establish the asymptotic normality, we define the linearization

$$\mathcal{L}_{nb}\left(\gamma_\pm\right)=M_n\left(\gamma_0,h_{0,b}\right)+\Upsilon_{1b,\pm}\left(\gamma_\pm-\gamma_0\right)+\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]. \qquad \text{(A.34)}$$

It is easy to see that the minimizer $\bar{\gamma}_\pm$ of $\left|\mathcal{L}_{nb}\left(\gamma_\pm\right)\right|^2$ satisfies

$$\sqrt{n/b}\left(\bar{\gamma}_\pm-\gamma_0\right)=-\Upsilon_{1b,\pm}^{-1}\sqrt{n/b}\left\{M_n\left(\gamma_0,h_{0,b}\right)+\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right\}. \qquad \text{(A.35)}$$

By conditions (B.2)-(B.5) and the $\sqrt{n/b}$-consistency of $\hat{\gamma}$,

$$\begin{aligned}&\left|M_n(\hat{\gamma}_\pm,\hat{h}_b)-\mathcal{L}_{nb}\left(\hat{\gamma}_\pm\right)\right| \\ =\;&\left|M_n(\gamma_0,h_{0,b})+M(\hat{\gamma}_\pm,\hat{h}_b)+M_n(\hat{\gamma}_\pm,\hat{h}_b)-M(\hat{\gamma}_\pm,\hat{h}_b)-M_n(\gamma_0,h_{0,b})-\mathcal{L}_{nb}\left(\hat{\gamma}_\pm\right)\right| \\ =\;&\left|M_n(\hat{\gamma}_\pm,\hat{h}_b)-M_n(\gamma_0,h_{0,b})-\Upsilon_{1b,\pm}\left(\hat{\gamma}-\gamma_0\right)-\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right| \\ \leq\;&\left|M_n(\hat{\gamma}_\pm,\hat{h}_b)-M(\hat{\gamma}_\pm,\hat{h}_b)-M_n(\gamma_0,h_{0,b})\right| \\ &+\left|M(\hat{\gamma}_\pm,h_{0,b})-\Upsilon_{1b,\pm}\left(\hat{\gamma}_\pm-\gamma_0\right)\right| \\ &+\left|M(\hat{\gamma}_\pm,\hat{h}_b)-M(\hat{\gamma}_\pm,h_{0,b})-\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right| \\ =\;&o_P\left((n/b)^{-1/2}\right)+o_P\left((n/b)^{-1/2}\right)+o_P\left((n/b)^{-1/2}\right)=o_P\left((n/b)^{-1/2}\right).\end{aligned}$$

By the same token and condition (B.6), $\left|M_n(\bar{\gamma}_\pm,\hat{h}_b)-\mathcal{L}_{nb}\left(\bar{\gamma}_\pm\right)\right|=o_P\left((n/b)^{-1/2}\right)$. Following the proof of Theorem 3.3 in Pakes and Pollard (1989), one can show that $\sqrt{n/b}\left(\hat{\gamma}_\pm-\bar{\gamma}_\pm\right)=o_P\left(1\right)$. It follows that

$$\begin{aligned}\sqrt{n/b}\left(\hat{\gamma}_\pm-\gamma_0\right) &= \sqrt{n/b}\left(\bar{\gamma}_\pm-\gamma_0\right)+o_P\left(1\right) \\ &= -\Upsilon_{1b,\pm}^{-1}\sqrt{n/b}\left\{M_n\left(\gamma_0,h_{0,b}\right)+\Upsilon_{2b}\left(\gamma_0,h_{0,b}\right)[\hat{h}_b-h_{0,b}]\right\}+o_P\left(1\right)\xrightarrow{d}N\left(0,\Omega_\gamma\right)\end{aligned}$$

where $\Omega_\gamma=\lim_{n\to\infty}\Upsilon_{1b,\pm}^{-1}V_\gamma\Upsilon_{1b,\pm}^{-1}$. ■

Proof of Theorem 3.2. We prove the theorem by verifying conditions (B.1)-(B.6) in Theorem A.7. (B.1) is satisfied by (2.20). Recall that $M(\gamma, h) = E\{[Y_t - \alpha(X_t, \gamma) - \beta(\gamma) D_t(\gamma)] w_t\}$, $h_{0,b} = (\alpha_{0,b}, \beta_{0,b})$, $\alpha_{0,b}(x,\gamma) = \alpha_0(x) + \delta_{\alpha,b}(x,\gamma)$, and $\beta_{0,b}(\gamma) = \beta_0 + \beta_0 c_{d_b}(\gamma)$. $\frac{d}{d\gamma_+}$ and $\frac{d}{d\gamma_-}$ denote the right and left derivative operators with respect to γ, respectively. Then

$$\begin{aligned}
\Upsilon_{1b,\pm}(\gamma, h_{0,b}) &= \frac{d}{d\gamma_\pm} M(\gamma, h_{0,b}) = -\frac{d}{d\gamma} E\{\delta_{\alpha,b}(X_t,\gamma) w(X_t) + \beta_0 c_{d_b}(\gamma) D_t(\gamma) w(X_t)\} \\
&= -\frac{d}{d\gamma_\pm}\left\{\int \beta_0 E\{K_b(X_t - x)\}[D_t(\gamma_0) - D_t(\gamma)] w(x)\, dx \right. \\
&\quad \left. -\beta_0 c_{d_b}(\gamma) \int E\{K_b(X_t - x)\} D_t(\gamma)] w(x)\, dx \ + \beta_0 c_{d_b}(\gamma) E[D_t(\gamma) w(X_t)]\right\} \\
&= \beta_0 \frac{d}{d\gamma_\pm}\left\{\int c_{0b}(x,\gamma) w(x)\, dx + b c_{d_b}(\gamma) \bar{c}_{0b}(\gamma)\right\}, \qquad \text{(A.36)}
\end{aligned}$$

where $\bar{c}_{0b}(\gamma) \equiv \frac{1}{b}[\int c_{0b}(x,\gamma) w(x)\, dx - c_w(\gamma)]$, $c_w(\gamma) \equiv E[D_t(\gamma) w(X_t)]$ and $c_{0b}(x,\gamma) \equiv E[K_b(X_t - x) D_t(\gamma)]$. In Appendix B, we derive the derivatives of $\int c_{0b}(x,\gamma) w(x)\, dx$, $c_w(\gamma)$ and $c_{d_b}(\gamma)$. In particular, we show that first two terms have continuous derivatives with respect to γ that are $O(1)$, the last term has continuous derivative at $\gamma \neq \gamma_0$ and both right and left continuous derivatives at $\gamma = \gamma_0$, and $b \cdot \frac{d}{d\gamma_\pm} c_{d_b}(\gamma)$ has a finite limit. In addition, $\frac{1}{b}[\int c_{0b}(x,\gamma) w(x)\, dx - c_w(\gamma)] = O(1)$. These facts imply that $\Upsilon_{1b,\pm}(\gamma, h_{0,b})$ is well behaved for each $\gamma \in \Gamma_\delta$ and $\Upsilon_{1b,\pm}(\gamma, h_{0,b})$ is continuous at $\gamma = \gamma_0$. (3.8) or (B.20) in Appendix B gives the the formula for $\Upsilon_{1b,\pm} \equiv \Upsilon_{1b,\pm}(\gamma_0, h_{0,b})$. Assumption A3(iii) ensures $\Upsilon_{1b,\pm}$ is bounded away from zero as $n \to \infty$. This verifies condition (B.2).

To verify condition (B.3), by direct calculation we have

$$\Upsilon_{2b}(\gamma, h)[h' - h] = -\int [\alpha'(x,\gamma) - \alpha(x,\gamma)] w(x)\, dF(x) - [\beta'(\gamma) - \beta(\gamma)] \int 1\{x_1 > \gamma\} w(x)\, dF(x). \qquad \text{(A.37)}$$

Noting that $M(\gamma, h)$ is linear in $h = (\alpha, \beta)$, condition (B.3(i)) is automatically satisfied for $c_{Mb} = 0$: $M(\gamma,\, h) - M(\gamma, h_{0,b}) - \Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}] = 0$. To check condition (B.3(ii)), observe that

$$\begin{aligned}
&\Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}] - \Upsilon_{2b}(\gamma_0, h_{0,b})[h - h_{0,b}] \\
= & -\left\{\int [\alpha(x,\gamma) - \alpha_{0,b}(x,\gamma)] w(x)\, dF(x) + [\beta(\gamma) - \beta_{0,b}(\gamma)] \int 1\{x_1 > \gamma\} w(x)\, dF(x)\right\} \\
& +\left\{\int [\alpha(x,\gamma_0) - \alpha_{0,b}(x,\gamma_0)] w(x)\, dF(x) + [\beta(\gamma_0) - \beta_{0,b}(\gamma_0)] \int 1\{x_1 > \gamma_0\} w(x)\, dF(x)\right\} \\
= & -\int [\alpha(x,\gamma) - \alpha_{0,b}(x,\gamma) - \alpha(x,\gamma_0) + \alpha_{0,b}(x,\gamma_0)] w(x)\, dF(x) \\
& -[\beta(\gamma) - \beta_{0,b}(\gamma) - \beta(\gamma_0) + \beta_{0,b}(\gamma_0)] \int 1\{x_1 > \gamma_0\} w(x)\, dF(x) \\
& -[\beta(\gamma) - \beta_{0,b}(\gamma)] \int 1\{x_1 > \gamma\} - 1\{x_1 > \gamma_0\} w(x)\, dF(x).
\end{aligned}$$

For all $h = (\alpha, \beta) \in \mathcal{H}_{\tilde{\delta}_{2n}}$ with $\tilde{\delta}_{2n} = o(b)$, we have $\sup_{x \in \mathcal{X}_0} |\alpha(x,\gamma) - \alpha_{0,b}(x,\gamma) - \alpha(x,\gamma_0) + \alpha_{0,b}(x,\gamma_0)|$ $= o(|\gamma - \gamma_0|)$ and $|\beta(\gamma) - \beta_{0,b}(\gamma) - \beta(\gamma_0) + \beta_{0,b}(\gamma_0)| = o(|\gamma - \gamma_0|)$. This implies that the first two terms on the rhs of the last equation is $o(|\gamma - \gamma_0|)$. The last term is also $o(|\gamma - \gamma_0|)$ because $\frac{1}{b}[\beta(\gamma) - \beta_{0,b}(\gamma)] = o(1)$ and $\int [1\{x_1 > \gamma\} - 1\{x_1 > \gamma_0\}] w(x)\, dF(x) = O(|\gamma - \gamma_0|)$. Hence $\Upsilon_{2b}(\gamma, h_{0,b})[h - h_{0,b}] - \Upsilon_{2b}(\gamma_0, h_{0,b})[h - h_{0,b}] = o(|\gamma - \gamma_0|)$.

To verify the first part of condition (B.4), we need to verify that $\sup_{x \in \mathcal{X}_0} |\hat{\alpha}_b(x,\gamma) - \alpha_{0,b}(x,\gamma) - \hat{\alpha}_b(x,\gamma_0) + \alpha_{0,b}(x,\gamma_0)| = o_P(|\gamma - \gamma_0|)$ and $\left|\hat{\beta}_b(\gamma) - \beta_{0,b}(\gamma) - \hat{\beta}_b(\gamma_0) + \beta_{0,b}(\gamma_0)\right| = o_P(|\gamma - \gamma_0|)$. We only outline the proof of the second part as the proof of the first part is analogous. By (A.4) and the fact that $\beta_{0,b}(\gamma) - \beta_{0,b}(\gamma_0) = \delta_{\beta,b}(\gamma)$, we have

$$\begin{aligned}
& \hat{\beta}_b(\gamma) - \beta_{0,b}(\gamma) - \hat{\beta}_b(\gamma_0) + \beta_{0,b}(\gamma_0) \\
= & \frac{1}{nb}\sum_{t=1}^{n}\left[S_{nb}^{-1}(\gamma)\tilde{D}_t(\gamma) - S_{nb}^{-1}(\gamma_0)\tilde{D}_t(\gamma_0)\right]\tilde{\varepsilon}_t \\
& + \frac{1}{nb}\sum_{t=1}^{n}\left[S_{nb}^{-1}(\gamma)\tilde{D}_t(\gamma) - S_{nb}^{-1}(\gamma_0)\tilde{D}_t(\gamma_0)\right]\tilde{\alpha}_0(X_t) \\
& + \left\{\beta_0 S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] - \delta_{\beta,b}(\gamma)\right\} \\
\equiv & v_{nb}(\gamma,\gamma_0) + b_{nb}(\gamma,\gamma_0) + r_{nb}(\gamma,\gamma_0), \text{ say,}
\end{aligned}$$

where $\tilde{\varepsilon}_t = n^{-1}\sum_{s=1}^{n} K_b(X_s - X_t)(\varepsilon_s - \varepsilon_t)$ and $\tilde{\alpha}_0(X_t) = n^{-1}\sum_{s=1}^{n} K_b(X_s - X_t)[\alpha_0(X_s) - \alpha_0(X_t)]$. We want to show that $v_{nb}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$, $b_{nb}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$, and $r_{nb}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$. We further decompose $v_{nb}(\gamma,\gamma_0)$ as follows:

$$\begin{aligned}
v_{nb}(\gamma,\gamma_0) &= \left[S_{nb}^{-1}(\gamma) - S_{nb}^{-1}(\gamma_0)\right]\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\tilde{\varepsilon}_t + \frac{1}{nb}S_{nb}^{-1}(\gamma_0)\sum_{t=1}^{n}\left[\tilde{D}_t(\gamma) - \tilde{D}_t(\gamma_0)\right]\tilde{\varepsilon}_t \\
&\equiv v_{nb,1}(\gamma,\gamma_0) + v_{nb,2}(\gamma,\gamma_0), \text{ say.}
\end{aligned}$$

Following the analysis of $\tilde{v}_{nb}(\gamma)$ in the proof of Lemma A.2, we can readily show $\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\tilde{\varepsilon}_t = O_P\left(n^{-1/2}b^{-1/2}\right)$ and $|S_{nb}(\gamma) - S_{nb}(\gamma_0)| = O_P(|\gamma - \gamma_0|/b)$. These, in conjunction with the fact that $S_{nb}(\gamma_0)$ converges in probability to a positive number, imply that $v_{nb,1}(\gamma,\gamma_0) = O_P\left(|\gamma - \gamma_0|\, n^{-1/2}b^{-3/2}\right) = o_P(|\gamma - \gamma_0|)$. Now

$$\begin{aligned}
S_{nb}(\gamma_0)\, v_{nb,2}(\gamma,\gamma_0) &= \frac{1}{n^2 b}\sum_{t=1}^{n}\sum_{s=1}^{n} K_b(X_s - X_t)[D_s(\gamma) - D_s(\gamma_0)]\tilde{\varepsilon}_t \\
&\quad - \frac{1}{n^2 b}\sum_{t=1}^{n}\sum_{s=1}^{n} K_b(X_s - X_t)[D_t(\gamma) - D_t(\gamma_0)]\tilde{\varepsilon}_t.
\end{aligned}$$

By straightforward moment calculations, we can bound each term on the right hand side (rhs) of the last expression by $o_P(|\gamma - \gamma_0|)$. It follows that $v_{nb,2}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$ and $v_{nb}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$. Analogously, we can show that $b_{nb}(\gamma,\gamma_0) = o_P(|\gamma - \gamma_0|)$. For $r_{nb}(\gamma,\gamma_0)$, we further make the following decomposition:

$$\begin{aligned}
r_{nb}(\gamma,\gamma_0) &= \beta_0 S_{nb}^{-1}(\gamma)\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] - \delta_{\beta,b}(\gamma) \\
&= \beta_0\left(S_{nb}^{-1}(\gamma) - \{E[S_{nb}(\gamma)]\}^{-1}\right)\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] \\
&\quad + \beta_0\left((E[S_{nb}(\gamma)])^{-1} - \left\{bE\left[d_b^2(X_t,\gamma)\right]\right\}^{-1}\right)\frac{1}{n}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] \\
&\quad + \beta_0\left\{bE\left[d_b^2(X_t,\gamma)\right]\right\}^{-1} \\
&\quad \times\left\{\frac{1}{n}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] - bE\{d_b(X_t,\gamma)[d_b(X_t,\gamma) - d_b(X_t,\gamma_0)]\}\right\} \\
&\equiv r_{nb,1}(\gamma,\gamma_0) + r_{nb,2}(\gamma,\gamma_0) + r_{nb,3}(\gamma,\gamma_0), \text{ say.}
\end{aligned}$$

By moment calculations and Chebyshev and Markov inequalities, we can show that $S_{nb}(\gamma) - E[S_{nb}(\gamma)] = O_P\left(n^{-1/2}b^{-1/2}\right) = o_P(b)$ and $\frac{1}{nb}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0) - \tilde{D}_t(\gamma)\right] = O_P(|\gamma-\gamma_0|/b)$. It follows that $r_{nb,1}(\gamma,\gamma_0) = o_P(|\gamma-\gamma_0|)$. By the same token, $r_{nb,2}(\gamma,\gamma_0) = o_P(|\gamma-\gamma_0|)$. In addition, we can show that

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)\right] - bE\left\{d_b(X_t,\gamma)\left[d_b(X_t,\gamma_0)-d_b(X_t,\gamma)\right]\right\}\\
=\;&\frac{1}{n}\sum_{t=1}^{n}\left[\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)\right] - E\left\{\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)\right]\right\}\right]\\
&+\frac{1}{n}\sum_{t=1}^{n}\left[E\left\{\tilde{D}_t(\gamma)\left[\tilde{D}_t(\gamma_0)-\tilde{D}_t(\gamma)\right]\right\} - bE\left\{d_b(X_t,\gamma)\left[d_b(X_t,\gamma_0)-d_b(X_t,\gamma)\right]\right\}\right]\\
=\;&o_P(|\gamma-\gamma_0|) + o(|\gamma-\gamma_0|) = o_P(|\gamma-\gamma_0|).
\end{aligned}$$

It follows that $r_{nb}(\gamma,\gamma_0) = o_P(|\gamma-\gamma_0|)$ and we have verified that $|\hat{\beta}_b(\gamma) - \beta_{0,b}(\gamma) - \hat{\beta}_b(\gamma_0) + \beta_{0,b}(\gamma_0)| = o_P(|\gamma-\gamma_0|)$.

To verify condition (B.6), observe that by (A.37),

$$\begin{aligned}
&\sqrt{n/b}\left\{M_n(\gamma_0,h_{0,b}) + \Upsilon_{2b}(\gamma_0,h_{0,b})[\hat{h}_b - h_{0,b}]\right\}\\
=\;&\frac{1}{\sqrt{nb}}\sum_{t=1}^{n}\varepsilon_t w(X_t) - \sqrt{\frac{n}{b}}\left\{\int[\hat{\alpha}_b(x,\gamma_0)-\alpha_0(x)]\,w(x)\,dF(x) + [\hat{\beta}_b(\gamma_0)-\beta_0]\int 1\{x_1>\gamma_0\}\,w(x)\,dF(x)\right\}.
\end{aligned}$$

By Lemma A.2(i) and the fact that $\psi_{1b}(\cdot,\gamma_0) = 0$, $\beta_{0,b}(\gamma_0) = \beta_0$, and that $b^{\upsilon} = o\left((n/b)^{-1/2}\right)$ under Assumption A4,

$$\hat{\beta}_b(\gamma_0) - \beta_0 = S_b^{-1}(\gamma_0)\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)]\,\varepsilon_t + o_P\left((n/b)^{-1/2}\right),$$

By Lemma A.3(ii) and the fact that $b^{\upsilon} = o\left((n/b)^{-1/2}\right)$ under Assumption A4,

$$\begin{aligned}
\hat{\alpha}_b(x,\gamma_0) - \alpha_0(x) \;=\;& f(x)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b(X_t-x)\,\varepsilon_t\\
&-f(x)^{-1}c_{0b}(x,\gamma)\,S_b^{-1}(\gamma_0)\frac{1}{n}\sum_{t=1}^{n}[c_{1b}(X_t,\gamma_0)-c_{2b}(X_t,\gamma_0)]\,\varepsilon_t + o_P\left((n/b)^{-1/2}\right)
\end{aligned}$$

uniformly in $x \in \mathcal{X}_0$. It follows that

$$\begin{aligned}
&\sqrt{n/b}\left\{M_n(\gamma_0,h_{0,b}) + \Upsilon_{2b}(\gamma_0,h_{0,b})[\hat{h}_b - h_{0,b}]\right\}\\
=\;&\frac{1}{\sqrt{nb}}\sum_{t=1}^{n}\left[w(X_t) - \int K_b(X_t-x)\,w(x)\,dx\right]\varepsilon_t + \bar{c}_{0b}(\gamma_0)\,S_b^{-1}(\gamma_0)\,\xi_n + o_P(1)\\
\equiv\;&A_{n,1} + A_{n,2} + o_P(1),\ \text{say},
\end{aligned}$$

where $\xi_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}b^{1/2}[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)]\varepsilon_t$, and $\bar{c}_{0b}(\gamma_0) = \frac{1}{b}\int[c_{0b}(x,\gamma_0) - 1\{x_1>\gamma_0\}f(x)]w(x)\,dx$. By straightforward calculations, we can verify that $E\left(A_{n,1}^2\right) = O\left(b^3\right)$ and

$$E\left(A_{n,2}^2\right) = \bar{c}_{0b}^2(\gamma_0)\,S_b^{-1}(\gamma_0)\,bE\left\{\sigma^2(X_t)\left[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)\right]^2\right\} \to V_{\gamma} > 0,$$

where $V_\gamma = \lim_{n\to\infty} V_{\gamma,b}$ and $V_{\gamma,b}$ is defined in (3.7). In addition, it is straightforward to verify that $\sum_{t=1}^n E\left|n^{-1/2}b^{1/2}[c_{1b}(X_t,\gamma_0) - c_{2b}(X_t,\gamma_0)]\varepsilon_t\right|^4 = O\left(n^{-1}b^{-2}\right) = o(1)$. It follows from the martingale central limit theorem that

$$\sqrt{n/b}\left\{M_n(\gamma_0,h_{0,b}) + \Upsilon_{2b}(\gamma_0,h_{0,b})[\hat{h}_b - h_{0,b}]\right\} \xrightarrow{D} N(0,V_\gamma).$$

By Theorem 3.1 and Lemma A.6, condition (B.5) is satisfied for $\varkappa = 0$. Then by the first part of Theorem A.7, $\hat{\gamma} - \gamma_0 = O_P\left(n^{-1/2}\log n\right) = o_P(n^{-\varkappa_0})$ for any $\varkappa_0 < 1/2$. With this result, we apply Lemma A.6 once more to conclude that condition (B.5) is satisfied for $\varkappa \in \left(\frac{3}{2}\eta, \frac{1}{2}\wedge(\kappa(1-\frac{\lambda}{2d})+\eta\right)$. Consequently, the second part of Theorem A.7 follows and we have

$$\begin{aligned}\sqrt{n/b}(\hat{\gamma}_\pm - \gamma_0) &= -\Upsilon_{1b,\pm}^{-1}\sqrt{n/b}\left\{M_n(\gamma_0,h_{0,b}) + \Upsilon_{2b}(\gamma_0,h_{0,b})[\hat{h}_b - h_{0,b}]\right\} + o_P(1) \\ &= -\bar{c}_{0b}(\gamma_0)S_b^{-1}(\gamma_0)\Upsilon_{1b,\pm}^{-1}\xi_n + o_P(1) \xrightarrow{d} N(0,\Omega_{\gamma,\pm}) \qquad \text{(A.38)}\end{aligned}$$

where $\Omega_{\gamma,\pm} = \lim_{n\to\infty}\Upsilon_{1b,\pm}^{-1}V_{\gamma,b}\Upsilon_{1b,\pm}^{-1}$. This completes the proof of the theorem. ■

Proof of Theorem 3.3. (i) Using the notations defined in the proof of Lemma A.2(i), we have

$$\hat{\beta}_b(\hat{\gamma}) = \beta_0 + v_{nb}(\gamma_0) + b_{nb}(\hat{\gamma}) + r_{nb}(\hat{\gamma}) + [v_{nb}(\hat{\gamma}) - v_{nb}(\gamma_0)]. \qquad \text{(A.39)}$$

By (A.10), $v_{nb}(\gamma) = S_b^{-1}(\gamma)v_{nb}^0(\gamma) + o_P((nb)^{-1/2})$ uniformly in $\gamma\in\Gamma$, where $v_{nb}^0(\gamma) = \frac{1}{n}\sum_{t=1}^n [c_{1b}(X_t,\gamma) - c_{2b}(X_t,\gamma)]\varepsilon_t$. Using the arguments as used in the proof of Lemma A.4(i), we can readily show that $\sup_{|\gamma-\gamma_0|\le C(n/b)^{-1/2}}\left|v_{nb}^0(\gamma) - v_{nb}^0(\gamma_0)\right| = o_P\left((nb)^{-1/2}\right)$, which, in conjunction with $|S_b(\gamma) - S_b(\gamma_0)| \le O(|\gamma-\gamma_0|)$ and Theorem 3.2, implies that

$$v_{nb}(\hat{\gamma}) - v_{nb}(\gamma_0) = o_P\left((nb)^{-1/2}\right). \qquad \text{(A.40)}$$

By the proof of Lemma A.1,

$$|b_{nb}(\hat{\gamma})| \le \sup_{\gamma\in\Gamma}|b_{nb}(\gamma)| = O_P(b^\upsilon) = o_P\left((nb)^{-1/2}\right). \qquad \text{(A.41)}$$

By Lemma A.3 and Theorem 3.2,

$$\begin{aligned} r_{nb}(\hat{\gamma}) &= \beta_0 S_b(\hat{\gamma})^{-1}\left\{\theta_{\psi_1}(\hat{\gamma}) + \frac{3}{n}\sum_{t=1}^n[\psi_{1b}(X_t;\hat{\gamma}) - \theta_{\psi_1}(\hat{\gamma})]\right\} + o_P((nb)^{-1/2}) \\ &= \beta_0 S_b(\gamma_0)^{-1}\theta_{\psi_1}(\hat{\gamma}) + \beta_0 S_b(\gamma_0)^{-1}\frac{3}{n}\sum_{t=1}^n[\psi_{1b}(X_t;\hat{\gamma}) - \theta_{\psi_1}(\hat{\gamma})] + o_P((nb)^{-1/2}).\end{aligned}$$

Using the fact that $\psi_{1b}(\cdot;\gamma_0) = \theta_{\psi_1}(\gamma_0) = 0$ and the stochastic equicontinuity argument, we can show that

$$\frac{1}{n}\sum_{t=1}^n[\psi_{1b}(X_t;\hat{\gamma}) - \theta_{\psi_1}(\hat{\gamma})] = \frac{3}{n}\sum_{t=1}^n[\psi_{1b}(X_t;\hat{\gamma}) - \theta_{\psi_1}(\hat{\gamma}) - \psi_{1b}(X_t;\gamma_0) + \theta_{\psi_1}(\gamma_0)] = o_P((nb)^{-1/2}).$$

Recall that $\hat{\gamma}$ is also denoted as $\hat{\gamma}_+$ if it is larger than γ_0 and $\hat{\gamma}_-$ otherwise. As demonstrated in Appendix B, the left and right derivatives, $\dot{\theta}_{\psi_1,+}(\gamma_0)$ and $\dot{\theta}_{\psi_1,-}(\gamma_0)$, of $\theta_{\psi_1}(\gamma)$ exists and are continuous at $\gamma = \gamma_0$. It follows that

$$\begin{aligned} r_{nb}(\hat{\gamma}_\pm) &= \beta_0 S_b(\gamma_0)^{-1}\left[\theta_{\psi_1}(\hat{\gamma}_\pm) - \theta_{\psi_1}(\gamma_0)\right] + o_P((nb)^{-1/2}) \\ &= \beta_0 S_b(\gamma_0)^{-1}\dot{\theta}_{\psi_1\pm}(\gamma_0)(\hat{\gamma}_\pm - \gamma_0) + o_P((nb)^{-1/2}). \qquad \text{(A.42)}\end{aligned}$$

Combining (A.39)-(A.42) and applying (A.10) and (A.38), we have

$$\begin{aligned}\sqrt{nb}\left(\hat{\beta}_b\left(\hat{\gamma}_{\pm}\right)-\beta_0\right) &= \sqrt{nb}v_{nb}\left(\gamma_0\right)+\beta_0 S_b\left(\gamma_0\right)^{-1}\left[b\dot{\theta}_{\psi_1,\pm}\left(\gamma_0\right)\right]\sqrt{n/b}(\hat{\gamma}_{\pm}-\gamma_0)+o_P\left(1\right) \\ &= S_b^{-1}\left(\gamma_0\right)\left[1-\beta_0 b\dot{\theta}_{\psi_1,\pm}\left(\gamma_0\right)\Upsilon_{1b,\pm}^{-1}\bar{c}_{0b}\left(\gamma_0\right)\right]\xi_n+o_P\left(1\right)\overset{D}{\to}N\left(0,\Omega_{\beta,\pm}\right), \quad (A.43)\end{aligned}$$

where $\Omega_{\beta,\pm}=\lim_{n\to\infty}\Omega_{n\beta,\pm}$ and $\Omega_{n\beta,\pm}$ is defined in (3.10).

(ii) By Lemmas A.1 and A.2 and their proofs, we can write $\hat{\alpha}_b\left(x,\hat{\gamma}\right)=\alpha_0\left(x\right)+V_{nb}\left(x\right)+B_{nb}\left(x\right)+R_{nb}(x;\hat{\beta}_b\left(\hat{\gamma}\right),\hat{\gamma})$, where

$$\begin{aligned}R_{nb}(x;\hat{\beta}_b\left(\hat{\gamma}\right),\hat{\gamma}) &= \beta_0\hat{f}_b\left(x\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b\left(X_t-x\right)\left[D_t\left(\gamma_0\right)-D_t\left(\hat{\gamma}\right)\right] \\ &\quad -[\hat{\beta}_b\left(\hat{\gamma}\right)-\beta_0]\hat{f}_b\left(x\right)^{-1}\frac{1}{n}\sum_{t=1}^{n}K_b\left(X_t-x\right)D_t\left(\hat{\gamma}\right) \\ &= \beta_0 f\left(x\right)^{-1}E\left\{K_b\left(X_t-x\right)\left[D_t\left(\gamma_0\right)-D_t\left(\gamma\right)\right]\right\}|_{\gamma=\hat{\gamma}} \\ &\quad -[\hat{\beta}_b\left(\hat{\gamma}\right)-\beta_0]f\left(x\right)^{-1}E\left[K_b\left(X_t-x\right)D_t\left(\gamma_0\right)\right]+o_P\left(n^{-1/2}b^{-d/2}\right).\end{aligned}$$

Let $\dot{c}_{0b}\left(x,\gamma_0\right)=\partial c_{0b}\left(x,\gamma_0\right)/\partial\gamma$. Noting that $E\left\{K_b\left(X_t-x\right)\left[D_t\left(\gamma_0\right)-D_t\left(\gamma\right)\right]\right\}|_{\gamma=\hat{\gamma}}=-[c_{0b}\left(x,\hat{\gamma}\right)-c_{0b}\left(x,\gamma_0\right)]=-\dot{c}_{0b}\left(x,\gamma_0\right)\left(\hat{\gamma}-\gamma_0\right)+o_P\left(|\hat{\gamma}-\gamma_0|\right)=O_P((nb)^{-1/2})$ and $\hat{\beta}_b\left(\hat{\gamma}\right)-\beta_0=O_P((nb)^{-1/2})$, $R_{nb}(x;\hat{\beta}_b\left(\hat{\gamma}\right),\hat{\gamma})$ is asymptotically negligible in the case where $d=1$ and is not otherwise. By standard arguments,

$$V_{nb}\left(x\right)=f\left(x\right)^{-1}\sum_{t=1}^{n}K_b\left(X_t-x\right)\varepsilon_t+o_P\left(n^{-1/2}b^{-d/2}\right)\text{ and }B_{nb}\left(x\right)=O_P\left(b^{\upsilon}\right)=o_P\left(n^{-1/2}b^{-d/2}\right),$$

Then by (A.38) and (A.43),

$$\begin{aligned}&\sqrt{nb^d}\left[\hat{\alpha}_b\left(x,\hat{\gamma}_{\pm}\right)-\alpha_0\left(x\right)\right] \\ =\;& f\left(x\right)^{-1}\frac{\sqrt{b^d}}{\sqrt{n}}\sum_{t=1}^{n}K_b\left(X_t-x\right)\varepsilon_t-\beta_0 f\left(x\right)^{-1}\dot{c}_{0b}\left(x,\gamma_0\right)\sqrt{nb^d}\left(\hat{\gamma}_{\pm}-\gamma_0\right) \\ & -f\left(x\right)^{-1}c_{0b}\left(x,\gamma_0\right)\sqrt{nb^d}\left[\hat{\beta}_b\left(\hat{\gamma}\right)-\beta_0\right]+o_P\left(1\right) \\ =\;& f\left(x\right)^{-1}\frac{\sqrt{b^d}}{\sqrt{n}}\sum_{t=1}^{n}K_b\left(X_t-x\right)\varepsilon_t-b^{(d-1)/2}f\left(x\right)^{-1}\beta_0\left[b\dot{c}_{0b}\left(x,\gamma_0\right)\right]\bar{c}_{0b}\left(\gamma_0\right)S_b^{-1}\left(\gamma_0\right)\xi_n \\ & -b^{(d-1)/2}f\left(x\right)^{-1}c_{0b}\left(x,\gamma_0\right)S_b^{-1}\left(\gamma_0\right)\left\{1-\beta_0 b\dot{\theta}_{\psi_1,\pm}\left(\gamma_0\right)\Upsilon_{1b,\pm}^{-1}\bar{c}_{0b}\left(\gamma_0\right)\right\}\xi_n+o_P\left(1\right) \\ =\;& f\left(x\right)^{-1}\frac{\sqrt{b^d}}{\sqrt{n}}\sum_{t=1}^{n}K_b\left(X_t-x\right)\varepsilon_t-b^{(d-1)/2}c_{\alpha,b,\pm}\left(x\right)\xi_n+o_P\left(1\right) \\ & \overset{d}{\to}N\left(0,f\left(x\right)^{-1}\sigma^2\left(x\right)\int K\left(u\right)^2du+\Delta_{\alpha,\pm}\left(x;d\right)\right),\end{aligned}$$

where $\Delta_{\alpha,\pm}\left(x;d\right)=\lim_{n\to\infty}\Delta_{\alpha,b,\pm}\left(x;d\right)$, and $\Delta_{\alpha,b,\pm}\left(x;d\right)$ and $c_{\alpha,b,\pm}\left(x\right)$ are defined in (3.11) and (3.12), respectively. This completes the proof of the theorem. ■

References

Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Bosq, D., 1998. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Lecture Notes in Statistics 110. Springer-Verlag.

Chan, K. S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregerssive model. *Annals of Statistics* 21, 520-533.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 6B (Chapter 76), 5549-5632. New York: Elsevier Science.

Chen, X., Linton, O., Van Keilegom, I., 2003. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71, 1591-1608.

Doukhan, P., Massart, P., Rio, E., 1995. Invariance principle for absolutely regular empirical processes. *Ann. Inst. Henri Poincaré– Probabilités et Staistiques* 31, 393-427.

Gozalo, P., Linton, Q., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* 104, 1-48.

Hahn, J., Ridder, G., 2013. Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica* 81, 315-340.

Hansen, B. E., 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24, 726-748.

Lee, A.J., 1990. *U-statistics: Theory and Practice.* Marcel Dekker, New York.

Li, Q., Racine, J., 2007. *Nonparametric Econometrics: Theory and Practice.* Princeton University Press: Princeton.

Mammen, E., Rothe, C., Schienle, M., 2012. Nonparametric regression with nonparametrically generated covariates. *Annals of Statistics* 40, 1132-1170.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis,* 17, 571-599.

Merlevède, F., M. Peilgrad, M., Rio, E., 2009. Bernstein inequality and moderate deviations under strong mixing conditions. *IMS collections: High Dimensional Probability V.*, 273-292.

Van der Vaart, A., Wellner, J. A., 1996. *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York.

Supplementary Material On

"Nonparametric Threshold Regression: Estimation and Inference"

Daniel J. Henderson,[a] Christopher F. Parmeter,[b] and Liangjun Su[c]

[a] *Department of Economics, University of Alabama, Tuscaloosa*

[b] *Department of Economics, University of Miami*

[c] *School of Economics, Singapore Management University*

THIS APPENDIX PROVIDES PROOFS FOR SOME TECHNICAL LEMMAS IN THE ABOVE PAPER.

B Calculation of Derivatives

B.1 Derivatives of $\theta_{\psi_1}(\gamma)$

Recall that $\theta_{\psi_1}(\gamma) = bE\{d_b(X_t,\gamma)[d_b(X_t,\gamma_0) - d_b(X_t,\gamma)]\} = A(\gamma,\gamma_0) - A(\gamma,\gamma)$, where $A(\gamma,\gamma') = bE[d_b(X_t,\gamma)\, d_b(X_t,\gamma')]$. We calculate the derivatives of $A(\gamma,\gamma_0)$ and $A(\gamma,\gamma)$ with respect to γ, respectively, by restricting our attention to the case $\gamma - \gamma_0 = o(b)$ and then evaluate the derivative at γ_0. Without loss of generality, we assume that $|\gamma - \gamma_0| < b$.

Let $\bar{k}(t) = \int_{-1}^{t} k(t)\, dt$. We first decompose $d_b(x,\gamma)$ as follows:

$$\begin{aligned}
d_b(x,\gamma) &= \frac{1}{b}\int K(u)[1\{x_1 + bu_1 > \gamma\} - 1\{x_1 > \gamma\}] f(x+bu)\, du \\
&= \frac{1}{b}\int K(u)[1\{x_1 \le \gamma\} - 1\{x_1 + bu_1 \le \gamma\}] f(x)\, du \\
&\quad + \frac{1}{b}\int K(u)[1\{x_1 \le \gamma\} - 1\{x_1 + bu_1 \le \gamma\}][f(x+bu) - f(x)]\, du \\
&= \frac{f(x)}{b}\left[1\{x_1 \le \gamma\} - \bar{k}\left(\frac{\gamma - x_1}{b}\right)\right] \\
&\quad + \frac{1}{b}\int K(u)[1\{x_1 \le \gamma\} - 1\{x_1 + bu_1 \le \gamma\}][f(x+bu) - f(x)]\, du \\
&\equiv d_{1b}(x,\gamma) + d_{2b}(x,\gamma), \text{ say.}
\end{aligned}$$

With this decomposition, we can write

$$\begin{aligned}
A(\gamma,\gamma') &= bE\{d_{1b}(X_t,\gamma)\, d_{1b}(X_t,\gamma')\} + bE\{d_{1b}(X_t,\gamma)\, d_{2b}(X_t,\gamma')\} \\
&\quad + bE\{d_{2b}(X_t,\gamma)\, d_{1b}(X_t,\gamma')\} + bE\{d_{2b}(X_t,\gamma)\, d_{2b}(X_t,\gamma')\} \\
&\equiv A_1(\gamma,\gamma') + A_2(\gamma,\gamma') + A_3(\gamma,\gamma') + A_4(\gamma,\gamma'), \text{ say.}
\end{aligned}$$

We focus on the study of $A_1(\gamma,\gamma_0)$ and $A_1(\gamma,\gamma)$ and comment on the other terms. Using the fact that $k(\cdot)$ has compact support $[-1,1]$ and thus $\bar{k}(s) = 0$ for any $s \le -1$ and $= 1$ for any $s \ge 1$, we can write

$$1\{x_1 \le \gamma\} - \bar{k}\left(\frac{\gamma - x_1}{b}\right) = 1\{-b \le x_1 - \gamma \le 0\} - \bar{k}\left(\frac{\gamma - x_1}{b}\right) 1\{-b \le x_1 - \gamma \le b\}. \quad \text{(B.1)}$$

Then

$$\begin{aligned}
A_1(\gamma,\gamma_0) &= \frac{1}{b}\int e(x_1)\,1\{-b\le x_1-\gamma\le 0\}\,1\{-b\le x_1-\gamma_0\le 0\}\,dx_1 \\
&\quad -\frac{1}{b}\int e(x_1)\,1\{-b\le x_1-\gamma\le 0\}\,\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)1\{-b\le x_1-\gamma_0\le b\}\,dx_1 \\
&\quad -\frac{1}{b}\int e(x_1)\,\bar{k}\left(\frac{\gamma-x_1}{b}\right)1\{-b\le x_1-\gamma\le b\}\,1\{-b\le x_1-\gamma_0\le 0\}\,dx_1 \\
&\quad +\frac{1}{b}\int e(x_1)\,\bar{k}\left(\frac{\gamma-x_1}{b}\right)1\{-b\le x_1-\gamma\le b\}\,\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)1\{-b\le x_1-\gamma_0\le b\}\,dx_1 \\
&\equiv A_{11}(\gamma,\gamma_0)-A_{12}(\gamma,\gamma_0)-A_{13}(\gamma,\gamma_0)+A_{14}(\gamma,\gamma_0),
\end{aligned}$$

where $e(x_1)=\int f(x)^3\,dx_{-1}$. Let $\tilde{e}(x_1-\gamma_0)=e(x_1)$. Then tedious calculations yield

$$\begin{aligned}
A_{11}(\gamma,\gamma_0) &= \frac{1}{b}\int \tilde{e}(x_1-\gamma_0)\,1\{-b+\gamma-\gamma_0\le x_1-\gamma_0\le \gamma-\gamma_0\}\,1\{-b\le x_1-\gamma_0\le 0\}\,dx_1 \\
&= \int \tilde{e}(vb)\,1\left\{-1+\frac{\gamma-\gamma_0}{b}\le v\le \frac{\gamma-\gamma_0}{b}\right\}1\{-1\le v\le 0\}\,dv \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,dv1\{\gamma>\gamma_0\}+\int_{-1}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,dv1\{\gamma\le\gamma_0\}, \\
A_{12}(\gamma,\gamma_0) &= \frac{1}{b}\int \tilde{e}(x_1-\gamma_0)\,\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)1\{-b+\gamma-\gamma_0\le x_1-\gamma_0\le \gamma-\gamma_0\}\,1\{-b\le x_1-\gamma_0\le b\}\,dx_1 \\
&= \int \tilde{e}(bv)\,\bar{k}(-v)\,1\left\{-1+\frac{\gamma-\gamma_0}{b}\le v\le \frac{\gamma-\gamma_0}{b}\right\}1\{-1\le v\le 1\}\,dv \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}(-v)\,dv1\{\gamma>\gamma_0\}+\int_{-1}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}(-v)\,dv1\{\gamma\le\gamma_0\}, \\
A_{13}(\gamma,\gamma_0) &= \frac{1}{b}\int \tilde{e}(x_1-\gamma_0)\,\bar{k}\left(\frac{\gamma_0-x_1}{b}+\frac{\gamma-\gamma_0}{b}\right)1\{-b+\gamma\le x_1\le b+\gamma\}\,1\{-b\le x_1-\gamma_0\le 0\}\,dx_1 \\
&= \int \tilde{e}(bv)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)1\left\{-1+\frac{\gamma-\gamma_0}{b}\le v\le 1+\frac{\gamma-\gamma_0}{b}\right\}1\{-1\le v\le 0\}\,dv \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv1\{\gamma>\gamma_0\}+\int_{-1}^{0}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv1\{\gamma\le\gamma_0\},
\end{aligned}$$

and

$$\begin{aligned}
A_{14}(\gamma,\gamma_0) &= \frac{1}{b}\int \tilde{e}(x_1-\gamma_0)\,\bar{k}\left(\frac{\gamma_0-x_1}{b}+\frac{\gamma-\gamma_0}{b}\right)\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)1\{-b+\gamma-\gamma_0\le x_1-\gamma_0\le b+\gamma-\gamma_0\} \\
&\quad \times 1\{-b\le x_1-\gamma_0\le b\}\,dx_1 \\
&= \int \tilde{e}(bv)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,1\left\{-1+\frac{\gamma-\gamma_0}{b}\le v\le 1+\frac{\gamma-\gamma_0}{b}\right\}1\{-1\le v\le 1\}\,dv \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{1}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv1\{\gamma>\gamma_0\} \\
&\quad +\int_{-1}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv1\{\gamma\le\gamma_0\}.
\end{aligned}$$

It follows that

$$
\begin{aligned}
A_1(\gamma,\gamma_0) &= \left\{\int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,dv-\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}(-v)\,dv-\int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv\right.\\
&\left.+\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv\right\}\times 1\{\gamma>\gamma_0\}\\
&+\left\{\int_{-1}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,dv-\int_{-1}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}(-v)\,dv-\int_{-1}^{0}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv\right.\\
&\left.+\int_{-1}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv\right\}\times 1\{\gamma\le\gamma_0\}. \qquad \text{(B.2)}
\end{aligned}
$$

When $\gamma>\gamma_0$, we apply Leibniz's formula and the fact that $\bar{k}(1)=1$ to obtain

$$
\begin{aligned}
\frac{\partial A_1(\gamma,\gamma_0)}{\partial\gamma} &= -\frac{1}{b}\tilde{e}(-b+\gamma-\gamma_0)-\frac{1}{b}\tilde{e}(\gamma-\gamma_0)\,\bar{k}\left(-\frac{\gamma-\gamma_0}{b}\right)+\frac{1}{b}\tilde{e}(-b+\gamma-\gamma_0)\,\bar{k}\left(1-\frac{\gamma-\gamma_0}{b}\right)\\
&+\frac{1}{b}\tilde{e}(-b+\gamma-\gamma_0)\,\bar{k}(1)-\frac{1}{b}\tilde{e}(-b+\gamma-\gamma_0)\,\bar{k}(1)\,\bar{k}\left(1-\frac{\gamma-\gamma_0}{b}\right)\\
&-\frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv+\frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv\\
&= -\frac{1}{b}\tilde{e}(\gamma-\gamma_0)\,\bar{k}\left(-\frac{\gamma-\gamma_0}{b}\right)-\frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv\\
&+\frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv \text{ which is } O(1/b).
\end{aligned}
$$

It follows that the **right** derivative of $A_1(\gamma,\gamma_0)$ with respect to γ at γ_0 is given by

$$
\begin{aligned}
\left.\frac{\partial A_1(\gamma,\gamma_0)}{\partial\gamma}\right|_{\gamma\to\gamma_0+} &= -\frac{1}{2b}\tilde{e}(0)-\frac{1}{b}\int_{-1}^{0}\tilde{e}(vb)\,k(-v)\,dv+\frac{1}{b}\int_{-1}^{1}\tilde{e}(vb)\,\bar{k}(-v)\,k(v)\,dv\\
&= -\frac{1}{2b}\tilde{e}(0)-\frac{1}{b}\int_{-1}^{0}\left[\tilde{e}(0)+\tilde{e}'(0)\,bv\right]k(-v)\,dv+\frac{1}{b}\int_{-1}^{1}\left[\tilde{e}(0)+\tilde{e}'(0)\,bv\right]\bar{k}(-v)\,k(-v)\,dv\\
&+o(1)\\
&= -\frac{1}{2b}e(\gamma_0)+e'(\gamma_0)\left(\int_{-1}^{1}v\bar{k}(-v)\,k(-v)\,dv-\int_{-1}^{0}vk(-v)\,dv\right)+o(1) \qquad \text{(B.3)}
\end{aligned}
$$

where we use the fact that $\int_{-1}^{0}k(-v)\,dv=\frac{1}{2}$ and $\int_{-1}^{1}\bar{k}(-v)\,k(-v)\,dv=\frac{1}{2}\bar{k}(v)^2\Big|_{v=-1}^{1}=\frac{1}{2}$.

When $\gamma<\gamma_0$, we apply Leibniz's formula and the fact that $\bar{k}(-1)=0$ to obtain

$$
\begin{aligned}
\frac{\partial A_1(\gamma,\gamma_0)}{\partial\gamma} &= \frac{1}{b}\tilde{e}(\gamma-\gamma_0)-\frac{1}{b}\tilde{e}(\gamma-\gamma_0)\,\bar{k}\left(-\frac{\gamma-\gamma_0}{b}\right)\\
&-\frac{1}{b}\int_{-1}^{0}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv+\frac{1}{b}\int_{-1}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv
\end{aligned}
$$

It follows that the **left** derivative of $A_1(\gamma,\gamma_0)$ with respect to γ at γ_0 is given by

$$\begin{aligned}
\left.\frac{\partial A_1(\gamma,\gamma_0)}{\partial\gamma}\right|_{\gamma\to\gamma_0-} &= \frac{1}{2b}\tilde e(0) - \frac{1}{b}\int_{-1}^{0}\tilde e(vb)\,k(-v)\,dv + \frac{1}{b}\int_{-1}^{1}\tilde e(vb)\,k(-v)\,\bar k(-v)\,dv \\
&= \frac{1}{2b}\tilde e(0) - \frac{1}{b}\int_{-1}^{0}\left[\tilde e(0)+\tilde e'(0)\,bv\right]k(-v)\,dv + \frac{1}{b}\int_{-1}^{1}\left[\tilde e(0)+\tilde e'(0)\,bv\right]k(-v)\,\bar k(-v)\,dv \\
&\quad + o(1) \\
&= \frac{1}{2b}e(\gamma_0) + e'(\gamma_0)\left(\int_{-1}^{1} v\bar k(-v)\,k(-v)\,dv - \int_{-1}^{0} vk(-v)\,dv\right) + o(1). \qquad \text{(B.4)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
A_{11}(\gamma,\gamma) &= \frac{1}{b}\int \tilde e(x_1-\gamma_0)\,1\{-b+\gamma-\gamma_0 \le x_1-\gamma_0 \le \gamma-\gamma_0\}\,dx_1 = \int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,dv, \\
A_{12}(\gamma,\gamma) &= \frac{1}{b}\int \tilde e(x_1-\gamma_0)\,\bar k\left(\frac{\gamma_0-x_1}{b}+\frac{\gamma-\gamma_0}{b}\right)1\{-b+\gamma-\gamma_0 \le x_1-\gamma_0 \le \gamma-\gamma_0\}\,dx_1 \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(bv)\,\bar k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv = A_{13}(\gamma,\gamma),\ \text{and} \\
A_{14}(\gamma,\gamma) &= \frac{1}{b}\int \tilde e(x_1-\gamma_0)\,\bar k^2\left(\frac{\gamma_0-x_1}{b}+\frac{\gamma-\gamma_0}{b}\right)1\{-b+\gamma-\gamma_0 \le x_1-\gamma_0 \le b+\gamma-\gamma_0\}\,dx_1 \\
&= \int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,\bar k^2\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv.
\end{aligned}$$

It follows that

$$A_1(\gamma,\gamma) = \int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,dv - 2\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(bv)\,\bar k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + \int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,\bar k^2\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \qquad \text{(B.5)}$$

and

$$\begin{aligned}
&\frac{dA_1(\gamma,\gamma)}{d\gamma} \\
=\ & \frac{1}{b}\tilde e(\gamma-\gamma_0) - \frac{1}{b}\tilde e(-b+\gamma-\gamma_0) - \frac{2}{b}\tilde e(\gamma-\gamma_0)\,\bar k(0) + \frac{2}{b}\tilde e(-b+\gamma-\gamma_0)\,\bar k(1) \\
& + \frac{1}{b}\tilde e(b+\gamma-\gamma_0)\,\bar k^2(-1) - \frac{1}{b}\tilde e(-b+\gamma-\gamma_0)\,\bar k^2(1) \\
& - \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(bv)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,\bar k\left(-v+\frac{\gamma-\gamma_0}{b}\right)k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \\
=\ & -\frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde e(bv)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}} \tilde e(vb)\,\bar k\left(-v+\frac{\gamma-\gamma_0}{b}\right)k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv.
\end{aligned}$$

Then the **total** derivative of $A_1(\gamma,\gamma)$ with respect to γ evaluated at γ_0 is given by

$$\begin{aligned}
\left.\frac{dA_1(\gamma,\gamma)}{d\gamma}\right|_{\gamma=\gamma_0} &= -\frac{2}{b}\int_{-1}^{0}\tilde e(bv)\,k(-v)\,dv + \frac{2}{b}\int_{-1}^{1}\tilde e(vb)\,\bar k(-v)\,k(-v)\,dv \\
&= -\frac{2}{b}\int_{-1}^{0}\left[\tilde e(0)+\tilde e'(0)\,bv\right]k(-v)\,dv + \frac{2}{b}\int_{-1}^{1}\left[\tilde e(0)+\tilde e'(0)\,bv\right]\bar k(-v)\,k(-v)\,dv + o(1) \\
&= 2e'(\gamma_0)\left[\int_{-1}^{1} v\bar k(-v)\,k(-v)\,dv - \int_{-1}^{0} vk(-v)\,dv\right] + o(1). \qquad \text{(B.6)}
\end{aligned}$$

Combining (B.3)-(B.6), we obtain the **right** derivative of $A_1(\gamma,\gamma_0) - A_1(\gamma,\gamma)$ with respect to γ evaluated at γ_0 as

$$\left.\frac{d\left[A_1(\gamma,\gamma_0) - A_1(\gamma,\gamma)\right]}{d\gamma}\right|_{\gamma=\gamma_{0+}} = -\frac{1}{2b}e(\gamma_0) - e'(\gamma_0)\left(\int_{-1}^{1} v\bar{k}(-v)\,k(-v)\,dv - \int_{-1}^{0} vk(-v)\,dv\right) + o(1)$$

and the **left** derivative of $A_1(\gamma,\gamma_0) - A_1(\gamma,\gamma)$ with respect to γ evaluated at γ_0 as

$$\left.\frac{d\left[A_1(\gamma,\gamma_0) - A_1(\gamma,\gamma)\right]}{d\gamma}\right|_{\gamma=\gamma_{0-}} = \frac{1}{2b}e(\gamma_0) - e'(\gamma_0)\left(\int_{-1}^{1} v\bar{k}(-v)\,k(-v)\,dv - \int_{-1}^{0} vk(-v)\,dv\right) + o(1).$$

Now, note that

$$\begin{aligned}
d_{2b}(x,\gamma) &= \frac{1}{b}\int K(u)\left[1\{x_1 \le \gamma\} - 1\{x_1 + bu_1 \le \gamma\}\right]\left[f(x+bu) - f(x)\right]du \\
&= \sum_{j=1}^{v} \frac{b^{j-1}}{j!}\frac{\partial^j f(x)}{\partial x_1^j}\int k(u_1)\,u_1^j\left[1\{x_1 \le \gamma\} - 1\{x_1 + bu_1 \le \gamma\}\right]du_1 + o\left(b^{v-1}\right) \\
&= \sum_{j=1}^{v} \frac{b^{j-1}}{j!}\frac{\partial^j f(x)}{\partial x_1^j}\left[\bar{k}_j(1)\,1\{x_1 \le \gamma\} - \bar{k}_j\left(\frac{\gamma - x_1}{b}\right)\right] + o\left(b^{v-1}\right)
\end{aligned}$$

where $\bar{k}_j(v) = \int_{-1}^{v} k(u_1)\,u_1^j du_1$. Following the analysis of $A_1(\gamma,\gamma_0) - A_1(\gamma,\gamma)$ above, we can readily show that the derivatives of $A_j(\gamma,\gamma_0) - A_j(\gamma,\gamma)$, $j = 2,3,4$, both exist and are $O(1)$ for all $\gamma \in \Gamma_\delta$ with $\gamma \ne \gamma_0$. It follows that when $\gamma > \gamma_0$,

$$\begin{aligned}
\dot{\theta}_{\psi_1}(\gamma) &= -\frac{1}{b}\tilde{e}(\gamma-\gamma_0)\,\bar{k}\left(-\frac{\gamma-\gamma_0}{b}\right) - \frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{0}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \\
&\quad + \frac{1}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv + \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(bv)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \\
&\quad - \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + O(1) \qquad \text{(B.7)}
\end{aligned}$$

and when $\gamma < \gamma_0$,

$$\begin{aligned}
\dot{\theta}_{\psi_1}(\gamma) &= \frac{1}{b}\tilde{e}(\gamma-\gamma_0) - \frac{1}{b}\tilde{e}(\gamma-\gamma_0)\,\bar{k}\left(-\frac{\gamma-\gamma_0}{b}\right) - \frac{1}{b}\int_{-1}^{0}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \\
&\quad + \frac{1}{b}\int_{-1}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)\bar{k}(-v)\,dv + \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}}\tilde{e}(bv)\,k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv \\
&\quad - \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}}\tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + O(1). \qquad \text{(B.8)}
\end{aligned}$$

At $\gamma = \gamma_0$, both the left and right derivatives exist and are continuous. Consequently, $\theta_{\psi_1}(\gamma)$ has continuous derivatives at all $\gamma \in \Gamma_{\delta_{1n}}$ with $\gamma \ne \gamma_0$; the **right** derivative $\dot{\theta}_{\psi_1,+}(\gamma)$ of $\theta_{\psi_1}(\gamma)$ at γ_0 satisfies

$$b \cdot \dot{\theta}_{\psi_1,+}(\gamma_0) = -\frac{1}{2}e(\gamma_0) + O(b), \qquad \text{(B.9)}$$

and the **left** derivative $\dot{\theta}_{\psi_1,+}(\gamma)$ of $\theta_{\psi_1}(\gamma)$ at γ_0 satisfies

$$b \cdot \dot{\theta}_{\psi_1,-}(\gamma_0) = \frac{1}{2}e(\gamma_0) + O(b). \qquad \text{(B.10)}$$

B.2 Derivatives of $c_{d_b}(\gamma)$

Recall that $c_{d_b}(\gamma) = \left(E\left[d_b^2(X_t,\gamma)\right]\right)^{-1} E\{d_b(X_t,\gamma)[d_b(X_t,\gamma_0) - d_b(X_t,\gamma)]\} = S_b^{-1}(\gamma)\,\theta_{\psi_1}(\gamma)$, where $S_b(\gamma) = b \cdot E\left[d_b^2(X_t,\gamma)\right] = \sum_{j=1}^{4} A_j(\gamma,\gamma)$, and $A_j(\gamma,\gamma')$'s are defined above. For any $\gamma \in \Gamma_\delta$, we have shown that the derivatives of $S_b(\gamma)$ exists, and so does $\theta_{\psi_1}(\gamma)$ for $\gamma \neq \gamma_0$. As a result,

$$\dot{c}_{d_b,\pm}(\gamma) \equiv \frac{dc_{d_b}(\gamma)}{d\gamma_\pm} = \frac{S_b(\gamma)\,\dot{\theta}_{\psi_1,\pm}(\gamma) - \theta_{\psi_1}(\gamma)\,\dot{S}_b(\gamma)}{S_b^2(\gamma)} \tag{B.11}$$

where

$$\begin{aligned}\dot{S}_b(\gamma) \quad &\equiv \quad \frac{dS_b(\gamma)}{d\gamma} = -\frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{\frac{\gamma-\gamma_0}{b}} \tilde{e}(bv)\,k\left(-v + \frac{\gamma-\gamma_0}{b}\right)dv \\ &\quad + \frac{2}{b}\int_{-1+\frac{\gamma-\gamma_0}{b}}^{1+\frac{\gamma-\gamma_0}{b}} \tilde{e}(vb)\,\bar{k}\left(-v+\frac{\gamma-\gamma_0}{b}\right)k\left(-v+\frac{\gamma-\gamma_0}{b}\right)dv + O(1),\end{aligned}$$

and $\dot{\theta}_{\psi_1,\pm}(\gamma)$ are defined in (B.7)-(B.10). Note that when $\gamma \neq \gamma_0$, $\dot{\theta}_{\psi_1,+}(\gamma)$ or $\dot{\theta}_{\psi_1,-}(\gamma)$ becomes the ordinary derivative $\dot{\theta}_{\psi_1}(\gamma)$. Observing that $S_b(\gamma) = O(1)$, $\theta_{\psi_1}(\gamma) = O(1)$, $\dot{\theta}_{\psi_1,\pm}(\gamma) = O(1/b)$, and $\dot{S}_b(\gamma) = O(1/b)$, we have $\dot{c}_{d_b,\pm}(\gamma) = O(1/b)$.

At $\gamma = \gamma_0$, both the left and right derivatives of $\theta_{\psi_1}(\gamma)$ exist and are continuous, and the derivative of $S_b(\gamma)$ exists and is continuous. By (B.9)-(B.10) and the fact that $\theta_{\psi_1}(\gamma_0) = 0$, the **right** derivative $\dot{c}_{d_b,+}(\gamma)$ of $c_{d_b}(\gamma)$ at γ_0 satisfies

$$b \cdot \dot{c}_{d_b,+}(\gamma_0) = \frac{S_b(\gamma_0)\left[b \cdot \dot{\theta}_{\psi_1,+}(\gamma_0)\right] - b\cdot\theta_{\psi_1}(\gamma_0)\,\dot{S}_b(\gamma_0)}{S_b^2(\gamma_0)} = -\frac{e(\gamma_0)}{2S_b(\gamma_0)} + O(b), \tag{B.12}$$

and the **left** derivative $\dot{c}_{d_b,-}(\gamma)$ of $c_{d_b}(\gamma)$ at γ_0 satisfies

$$b \cdot \dot{c}_{d_b,-}(\gamma_0) = \frac{S_b(\gamma_0)\left[b \cdot \dot{\theta}_{\psi_1,-}(\gamma_0)\right] - b\cdot\theta_{\psi_1}(\gamma_0)\,\dot{S}_b(\gamma_0)}{S_b(\gamma_0)^2} = \frac{e(\gamma_0)}{2S_b(\gamma_0)} + O(b). \tag{B.13}$$

B.3 Derivatives of $c_{0b}(x,\gamma)$, $\int c_{0b}(x,\gamma)\,w(x)\,dx$, and $c_w(\gamma)$ with respect to γ

Recall that $c_{0b}(x,\gamma) = E[K_b(X_t - x)\,D_t(\gamma)] = \int K(u)\,1\left\{u_1 > \frac{\gamma-x_1}{b}\right\} f(x+bu)\,du$ and the univariate kernel function $k(\cdot)$ has compact support $[-1,1]$. We first make two observations: 1) If $x_1 \leq \gamma - b$, then $\frac{\gamma-x_1}{b} \geq 1$ and $c_{0b}(x,\gamma) = \int K(u)\cdot 0 \cdot f(x+bu)\,du = 0$; 2) If $x_1 \geq \gamma + b$, then $\frac{\gamma-x_1}{b} \leq -1$ and $c_{0b}(x,\gamma) = \int K(u)\,f(x+bu)\,du = f(x) + O(b^v)$. In either case, we have $\dot{c}_{0b}(x,\gamma) \equiv \partial c_{0b}(x,\gamma)/\partial\gamma = 0$. Below we focus on the calculation of the partial derivative for the case where $\gamma - b < x_1 < \gamma + b$.

Assuming that $x_1 \in (\gamma - b, \gamma + b)$, we have

$$\begin{aligned} c_{0b}(x,\gamma) \quad &= \quad \int k(u_1)\,1\{x_1 + bu_1 > \gamma\}\,f(x)\,du_1 + \int K(u)\,1\{x_1 + bu_1 > \gamma\}\,[f(x+bu) - f(x)]\,du \\ &= \quad f(x)\int k(u_1)\,1\left\{u_1 > \frac{\gamma-x_1}{b}\right\}du_1 + \sum_{j=1}^{v}\frac{b^j}{j!}\frac{\partial^j f(x)}{\partial x_1^j}\int k(u_1)\,u_1^j 1\left\{u_1 > \frac{\gamma-x_1}{b}\right\}du_1 + o(b^v) \\ &= \quad f(x)\left[1 - \bar{k}\left(\frac{\gamma-x_1}{b}\right)\right] + \sum_{j=1}^{v}\frac{b^j}{j!}\frac{\partial^j f(x)}{\partial x_1^j}\int_{\frac{\gamma-x_1}{b}}^{1} k(u_1)\,u_1^j\,du_1 + o(b^v).\end{aligned}$$

With this, one can readily show that when $x_1 \in (\gamma - b, \gamma + b)$,

$$\dot{c}_{0b}(x,\gamma) = -\frac{1}{b}f(x)\,k\left(\frac{\gamma-x_1}{b}\right) - \sum_{j=1}^{v}\frac{b^{j-1}}{j!}\frac{\partial^j f(x)}{\partial x_1^j}k\left(\frac{\gamma-x_1}{b}\right)\left(\frac{\gamma-x_1}{b}\right)^j + o(b^v),$$

and the partial derivative of $\int c_{0b}(x,\gamma)\,w(x)\,dx$ with respect to γ is given by

$$\begin{aligned}\int \dot{c}_{0b}(x,\gamma)\,w(x)\,dx &= -\frac{1}{b}\int\int_{\gamma-b}^{\gamma+b} f(x)\,k\left(\frac{\gamma-x_1}{b}\right)w(x)\,dx_1dx_{-1}+O(b)\\ &= -\int f(\gamma,x_{-1})\,w(\gamma,x_{-1})\,dx_{-1}+O(b)=-e_w(\gamma)+O(b). \qquad \text{(B.14)}\end{aligned}$$

where $e_w(x_1)=\int w(x)\,f(x)\,dx_{-1}$.

Now, recall that $c_w(\gamma)=E[D_t(\gamma)\,w(X_t)]$. Noting that $c_w(\gamma)=\int 1\{x_1>\gamma\}\,w(x)\,f(x)\,dx=\int_\gamma^\infty e_w(x_1)\,dx_1$, we have

$$\dot{c}_w(\gamma)=\frac{\partial c_w(\gamma)}{\partial\gamma}=-e_w(\gamma). \qquad \text{(B.15)}$$

B.4 Derivatives of $M(\gamma,h_{0b})$ with respect to γ

By (A.36) and the chain rule, we have

$$\begin{aligned}\Upsilon_{1b,\pm}(\gamma,h_{0,b}) &= \beta_0\int\dot{c}_{0b}(x,\gamma)\,w(x)\,dx+\beta_0 c_{d_b}(\gamma)\left[\int\dot{c}_{0b}(x,\gamma)\,w(x)\,dx-\dot{c}_w(\gamma)\right]\\ &\quad+\beta_0\left[b\cdot\dot{c}_{d_b,\pm}(\gamma)\right]\bar{c}_{0b}(\gamma), \qquad \text{(B.16)}\end{aligned}$$

where $\dot{c}_{d_b,\pm}(\gamma)$, $\int\dot{c}_{0b}(x,\gamma)\,w(x)\,dx$, and $\dot{c}_w(\gamma)$ are given in (B.11), (B.14), and (B.15), respectively. When $\gamma=\gamma_0$, we have

$$\Upsilon_{1b,\pm}(\gamma_0,h_{0,b})=-\beta_0 e_w(\gamma_0)\mp\beta_0\frac{e(\gamma_0)}{2S_b(\gamma_0)}\bar{c}_{0b}(\gamma_0)+O(b), \qquad \text{(B.17)}$$

where recall that $e_w(x_1)=\int w(x)\,f(x)\,dx_{-1}$, $e(x_1)=\int f(x)^3\,dx_{-1}$, $c_{0b}(x,\gamma)=E[K_b(X_t-x)\,D_t(\gamma)]$, and $c_w(\gamma)=E[D_t(\gamma)\,w(X_t)]$, and we have used the fact that $c_{d_b}(\gamma_0)=0$. Noting that $c_w(\gamma_0)=\int 1\{x_1>\gamma_0\}\,e_w(x_1)\,dx_1$ and that

$$\begin{aligned}\frac{1}{b}\int c_{0b}(x,\gamma_0)\,w(x)\,dx &= \frac{1}{b}\left\{\int\int k(u_1)\,1\left\{u_1>\frac{\gamma_0-x_1}{b}\right\}du_1 f(x)\,w(x)\,dx\right\}+O(b)\\ &= \frac{1}{b}\left\{\int\left[1-\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)\right]e_w(x_1)\,dx_1\right\}+O(b),\end{aligned}$$

and using (B.1) and change of variables, we have

$$\begin{aligned}\bar{c}_{0b}(\gamma_0) &= \frac{1}{b}\left[\int c_{0b}(x,\gamma_0)\,w(x)\,dx-c_w(\gamma_0)\right]\\ &= \frac{1}{b}\int e_w(x_1)\left[1\{x_1\le\gamma_0\}-\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)\right]dx_1+O(b)\\ &= \frac{1}{b}\int_{\gamma_0-b}^{\gamma_0}e_w(x_1)\,dx_1-\frac{1}{b}\int_{\gamma_0-b}^{\gamma_0+b}e_w(x_1)\,\bar{k}\left(\frac{\gamma_0-x_1}{b}\right)dx_1+O(b)\\ &= \frac{1}{b}\int_{\gamma_0-b}^{\gamma_0}e_w(x_1)\,dx_1-\int_{-1}^{1}e_w(\gamma_0+bv)\,\bar{k}(-v)\,dv+O(b)=\frac{1}{b}\int_{\gamma_0-b}^{\gamma_0}\left[e_w(x_1)-e_w(\gamma_0)\right]dx_1+O(b)\\ &= \frac{1}{b}\int_{\gamma_0-b}^{\gamma_0}\dot{e}_w(\gamma_0)(x_1-\gamma_0)\,dx_1+O(b)=-\frac{1}{2}\dot{e}_w(\gamma_0)+O(b), \qquad \text{(B.18)}\end{aligned}$$

where $\dot{e}_w(x_1) = \partial e_w(x_1)/\partial x_1$, and we have used the fact that $\int_{-1}^{1} \bar{k}(-v)\,dv = \int_{-1}^{1} \bar{k}(v)\,dv = 1$ by integration by parts and Assumption A3. By (B.5) and the fact that $\sum_{j=2}^{4} A(\gamma,\gamma) = O(b)$, we have

$$\begin{aligned} S_b(\gamma_0) &= A_1(\gamma_0,\gamma_0) + O(b) \\ &= \int_{-1}^{0} \tilde{e}(vb)\,dv - 2\int_{-1}^{0} \tilde{e}(bv)\,\bar{k}(-v)\,dv + \int_{-1}^{1} \tilde{e}(vb)\,\bar{k}^2(-v)\,dv + O(b) \\ &= \tilde{e}(0)\left[1 - 2\int_{0}^{1} \bar{k}(v)\,dv + \int_{-1}^{1} \bar{k}^2(v)\,dv\right] + O(b) = e(\gamma_0)\,c_{\bar{k}} + O(b) \end{aligned} \tag{B.19}$$

where $c_{\bar{k}} = 1 - 2\int_0^1 \bar{k}(v)\,dv + \int_{-1}^{1} \bar{k}^2(v)\,dv > 0$. Consequently,

$$\begin{aligned} \Upsilon_{1b,\pm}(\gamma_0, h_{0,b}) &= -\beta_0 e_w(\gamma_0) \pm \beta_0 \frac{e(\gamma_0)}{4S_b(\gamma_0)}\dot{e}_w(\gamma_0) + O(b) \\ &= -\beta_0 e_w(\gamma_0) \pm \beta_0 \frac{1}{4c_{\bar{k}}}\dot{e}_w(\gamma_0) + O(b). \end{aligned} \tag{B.20}$$

C Additional Simulation Results

Here we provide additional results on the performance of our semiparametric threshold estimator. Tables 8 to 11 provide bias and WASE results for DGPs 2-7, all using a signal to noise ratio of 0.75. Several key features emerge regardless of DGP, as n increases the bias of both $\widehat{\beta}$ and $\widehat{\gamma}$ decreases, the WASE for all three estimators decrease as the sample size increases, with the rate of decrease for $\widehat{\gamma}$ faster, as expected, than $\widehat{\beta}$ and $\widehat{\alpha}(x)$.

Table 6: Simulation Performance of Semiparametric Threshold Estimator, DGP 2, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-1.600	4.551	0.585	3.053	0.259
$n = 200$	-0.977	2.467	0.442	2.386	0.158
$n = 400$	-0.311	0.815	0.181	0.974	0.086
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-1.475	5.517	0.245	1.935	0.284
$n = 200$	-0.610	2.171	0.152	1.102	0.164
$n = 400$	-0.038	0.296	0.005	0.109	0.084
	$\beta = 2, \gamma = -1$				
$n = 100$	-1.148	5.338	0.146	1.279	0.300
$n = 200$	-0.258	1.299	0.014	0.315	0.163
$n = 400$	0.014	0.150	0.005	0.023	0.086
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.184	3.381	0.020	1.376	0.276
$n = 200$	0.124	1.422	0.006	0.808	0.157
$n = 400$	0.178	0.520	-0.012	0.362	0.085
	$\beta = 1.5, \gamma = 0$				
$n = 100$	0.339	2.910	-0.012	0.622	0.281
$n = 200$	0.369	1.065	-0.001	0.287	0.154
$n = 400$	0.337	0.315	0.009	0.044	0.085
	$\beta = 2, \gamma = 0$				
$n = 100$	0.413	2.951	0.013	0.457	0.286
$n = 200$	0.444	1.023	0.006	0.143	0.158
$n = 400$	0.334	0.305	-0.001	0.013	0.086

Table 7: Simulation Performance of Semiparametric Threshold Estimator, DGP 3, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-0.209	0.996	0.211	0.272	0.032
$n = 200$	0.093	0.264	0.042	0.050	0.017
$n = 400$	0.110	0.075	0.007	0.008	0.009
	$\beta = 1.5, \gamma = -1$				
$n = 100$	0.120	0.366	0.021	0.037	0.026
$n = 200$	0.151	0.106	0.003	0.005	0.014
$n = 400$	0.117	0.040	0.000	0.000	0.008
	$\beta = 2, \gamma = -1$				
$n = 100$	0.218	0.145	-0.006	0.002	0.024
$n = 200$	0.151	0.070	-0.002	0.000	0.014
$n = 400$	0.115	0.037	0.000	0.000	0.008
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.889	1.212	0.000	0.884	0.022
$n = 200$	-0.536	0.512	-0.002	0.313	0.013
$n = 400$	-0.314	0.157	0.001	0.037	0.007
	$\beta = 1.5, \gamma = 0$				
$n = 100$	-0.560	0.429	-0.008	0.024	0.015
$n = 200$	-0.395	0.184	-0.002	0.000	0.009
$n = 400$	-0.283	0.096	0.000	0.000	0.006
	$\beta = 2, \gamma = 0$				
$n = 100$	-0.532	0.314	-0.008	0.001	0.012
$n = 200$	-0.400	0.177	-0.002	0.000	0.008
$n = 400$	-0.283	0.090	0.000	0.000	0.005

Table 8: Simulation Performance of Semiparametric Threshold Estimator, DGP 4, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-0.764	2.935	0.795	1.499	0.118
$n = 200$	-0.921	2.626	0.757	1.144	0.078
$n = 400$	-1.016	2.484	0.781	1.015	0.055
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-0.807	3.664	0.557	1.059	0.138
$n = 200$	-0.954	3.107	0.499	0.794	0.093
$n = 400$	-0.854	2.405	0.439	0.594	0.065
	$\beta = 2, \gamma = -1$				
$n = 100$	-0.744	4.048	0.394	0.795	0.144
$n = 200$	-0.710	2.599	0.296	0.514	0.097
$n = 400$	-0.591	1.806	0.181	0.237	0.067
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.088	2.028	-0.058	0.679	0.110
$n = 200$	-0.074	1.402	-0.072	0.321	0.076
$n = 400$	-0.093	1.080	-0.089	0.132	0.054
	$\beta = 1.5, \gamma = 0$				
$n = 100$	-0.106	2.183	-0.084	0.461	0.124
$n = 200$	-0.073	1.306	-0.078	0.178	0.088
$n = 400$	0.013	0.645	-0.057	0.042	0.065
	$\beta = 2, \gamma = 0$				
$n = 100$	-0.319	2.869	-0.023	0.414	0.145
$n = 200$	-0.045	1.132	-0.049	0.107	0.099
$n = 400$	-0.006	0.556	-0.034	0.031	0.070

Table 9: Simulation Performance of Semiparametric Threshold Estimator, DGP 5, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-1.078	9.865	1.058	2.903	0.678
$n = 200$	-0.924	6.452	1.038	3.000	0.511
$n = 400$	-0.784	3.871	0.981	3.005	0.357
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-1.568	11.333	1.200	3.130	0.674
$n = 200$	-1.445	7.670	0.990	2.980	0.512
$n = 400$	-1.381	5.253	1.015	3.066	0.361
	$\beta = 2, \gamma = -1$				
$n = 100$	-2.021	13.251	1.264	3.353	0.683
$n = 200$	-2.029	9.981	1.082	3.189	0.524
$n = 400$	-2.059	7.818	0.884	2.906	0.374
	$\beta = 1, \gamma = 0$				
$n = 100$	0.150	10.700	-0.124	0.988	0.806
$n = 200$	0.022	6.893	-0.102	0.900	0.606
$n = 400$	0.310	3.448	-0.045	0.577	0.413
	$\beta = 1.5, \gamma = 0$				
$n = 100$	0.524	10.894	-0.108	0.620	0.839
$n = 200$	0.641	6.019	-0.066	0.501	0.620
$n = 400$	0.895	2.520	-0.042	0.212	0.410
	$\beta = 2, \gamma = 0$				
$n = 100$	0.659	10.603	-0.067	0.500	0.854
$n = 200$	1.000	5.332	-0.016	0.211	0.623
$n = 400$	0.982	2.374	0.009	0.065	0.411

Table 10: Simulation Performance of Semiparametric Threshold Estimator, DGP 6, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
	$\beta = 1, \gamma = -1$				
$n = 100$	-0.281	2.653	0.363	2.530	0.487
$n = 200$	-0.064	1.188	0.216	1.510	0.393
$n = 400$	0.091	0.348	0.077	0.518	0.286
	$\beta = 1.5, \gamma = -1$				
$n = 100$	-0.091	2.676	0.206	1.590	0.519
$n = 200$	0.032	1.123	0.082	0.766	0.397
$n = 400$	0.167	0.178	0.005	0.062	0.282
	$\beta = 2, \gamma = -1$				
$n = 100$	0.167	2.157	0.079	0.739	0.534
$n = 200$	0.177	0.726	0.053	0.379	0.402
$n = 400$	0.175	0.161	-0.005	0.005	0.289
	$\beta = 1, \gamma = 0$				
$n = 100$	-0.765	3.091	-0.053	2.988	0.488
$n = 200$	-0.566	1.696	0.016	2.224	0.393
$n = 400$	-0.235	0.583	-0.020	0.911	0.291
	$\beta = 1.5, \gamma = 0$				
$n = 100$	-0.763	3.353	-0.039	2.123	0.519
$n = 200$	-0.373	1.431	-0.015	1.212	0.397
$n = 400$	-0.034	0.254	0.002	0.181	0.289
	$\beta = 2, \gamma = 0$				
$n = 100$	-0.761	3.756	-0.016	1.534	0.537
$n = 200$	-0.232	1.198	-0.028	0.570	0.408
$n = 400$	-0.004	0.183	0.001	0.051	0.290

Table 11: Simulation Performance of Semiparametric Threshold Estimator, DGP 7, signal to noise ratio=0.75, 1000 Simulations

	β		γ		$\alpha(x)$
	Bias	MSE	Bias	MSE	WASE
			$\beta = 1, \gamma = -1$		
$n = 100$	-0.337	1.400	0.397	1.907	0.144
$n = 200$	-0.179	0.735	0.258	1.228	0.089
$n = 400$	-0.095	0.353	0.146	0.567	0.052
			$\beta = 1.5, \gamma = -1$		
$n = 100$	-0.297	1.670	0.296	1.202	0.161
$n = 200$	-0.185	0.835	0.152	0.634	0.090
$n = 400$	-0.037	0.249	0.034	0.119	0.052
			$\beta = 2, \gamma = -1$		
$n = 100$	-0.288	1.948	0.202	0.798	0.160
$n = 200$	-0.101	0.651	0.061	0.219	0.094
$n = 400$	-0.023	0.210	0.008	0.033	0.055
			$\beta = 1, \gamma = 0$		
$n = 100$	-0.367	1.425	-0.083	1.849	0.136
$n = 200$	-0.210	0.718	-0.130	1.073	0.085
$n = 400$	-0.105	0.301	-0.030	0.426	0.051
			$\beta = 1.5, \gamma = 0$		
$n = 100$	-0.282	1.487	-0.142	1.073	0.152
$n = 200$	-0.115	0.598	-0.038	0.372	0.088
$n = 400$	-0.001	0.170	-0.012	0.068	0.051
			$\beta = 2, \gamma = 0$		
$n = 100$	-0.296	1.683	-0.038	0.661	0.156
$n = 200$	-0.044	0.474	-0.016	0.141	0.089
$n = 400$	-0.012	0.217	-0.003	0.014	0.053