EFFICIENT KERNEL-BASED SEMIPARAMETRIC IV ESTIMATION WITH AN APPLICATION TO RESOLVING A PUZZLE ON THE ESTIMATES OF THE RETURN TO SCHOOLING

Feng Yao
Department of Economics
West Virginia University
Morgantown, WV 26505 USA
email: feng.yao@mail.wvu.edu
Voice: + 1 304 293 7867

Junsen Zhang
Department of Economics
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong
email: jszhang@cuhk.edu.hk
Voice: + 852 2609 8186

April, 2012

Abstract. An interesting puzzle in estimating the effect of education on labor market earnings (Card (2001)) is that the 2SLS estimate for the return to schooling typically exceeds the OLS estimate, but the 2SLS estimate is fairly imprecise. We provide a new explanation that it could be due to the restrictive linear functional form specification on the control variables and the reduced form. For the parameters of endogenous regressors, we propose three kernel-based semiparametric IV estimators that relax the tight functional form assumption on the control variables and the reduced form. They have explicit algebraic structures and are easily implemented without numerical optimizations. We show that these estimators are consistent, asymptotically normally distributed, and reach the semiparametric efficiency bound. A Monte Carlo study demonstrates that our estimators perform well in finite samples. We apply the proposed estimators to estimate the return to schooling in Card (1995). We find that the semiparametric estimates of the return to schooling are much smaller and more precise than the 2SLS estimate, and the difference largely comes from the misspecification in the linear reduced form.

Keywords and Phrases. Instrumental variables, semiparametric regression, efficient estimation.

JEL Classifications. C14, C21

1 Introduction

Many studies have been devoted to uncover the causal effect of education on labor market earnings, where endogeneity in education calls for the use of instrumental variables (IV) to estimate the return to schooling (see Card (2001) for an excellent summary). Let's consider the framework studied by Card (1995), $Y_t = Z_{1t}\alpha + X_t\beta + \epsilon_t$, where Y_t is the log of wages, X_t is the years of schooling, Z_{1t} includes a set of exogenous control variables and ϵ_t is the error term. One can conveniently interpret β as the return to schooling, but the ordinary least squares (OLS) estimate is not reliable since schooling is not randomly assigned. Various interesting studies summarized by Card (2001) utilize the supply-side variables, such as minimum school-leaving age, tuition costs, or the geographic proximity of schools, as instrumental variables to perform an IV type estimation, typically a two-stage least squares (2SLS) estimation.

An interesting finding from the studies (Card (2001), p1155) is that the IV estimate of β typically exceeds the corresponding OLS estimate, often by 20 percent or more, though the IV estimate is relatively imprecise. Assuming negative correlation between the omitted ability variables and the marginal cost of schooling, OLS methods should lead to upward-biased estimates of the true casual effect of schooling, and thus the even larger IV estimate presents a puzzle. One typical interpretation (Griliches (1977) and Angrist and Krueger (1991)) is that the OLS ability biases are relatively small, and the gaps between OLS and IV estimates are explained by the OLS estimate's downward bias largely due to measurement error[1]. Because the IV estimates are fairly imprecise, it is difficult to accept or reject this explanation in general. However, Card points out that measurement error bias can explain only 10 percent of the gap between OLS and IV. Thus, it is unlikely that many studies (see Table II in Card) find large positive gaps simply because of measurement error. Furthermore, neither Card nor other related studies analyze the cause of the imprecision of the IV estimates.

We provide an alternative explanation. The result that 2SLS provides a larger estimate of the

[1] We cite one for illustration. Card (2001) gives further explanations, which does not include our proposed explanation.

return to schooling than OLS could be due to restrictive linear functional forms placed on the control variables Z_{1t} and on the reduced form of X_t. Misspecification could lead to different and misleading estimates on return to schooling. We propose three kernel-based efficient semiparametric IV estimators that relax the restrictive functional form assumptions on the control variables and on the reduced form. We demonstrate in our empirical illustration section that the semiparametric IV estimates are much smaller than the 2SLS estimates, and they improve greatly on the precision. We thus conclude that one alternative explanation for the gap between 2SLS and OLS estimates on return to schooling is due to the restrictive functional form specification.

In this paper, we consider the following semiparametric additive regression model[2]

$$Y_t = g(Z_{1t}, X_t, \epsilon_t) = m(Z_{1t}) + X_t\beta + \epsilon_t, \qquad t = 1, \cdots, n. \tag{1}$$

We denote the endogenous explanatory variables explicitly by $X_t' = (X_{t,1}, \cdots, X_{t,K})' \in \Re^K$, the included exogenous variable by $Z_{1t}' = (Z_{1t,1}, \cdots, Z_{1t,l_1})' \in \Re^{l_1}$, and the excluded exogenous variable by $Z_{2t}' \in \Re^{l_2}$, with $l_1 + l_2 = l$. We assume the exogenous variables enter this equation of the model via a nonparametric function $m(\cdot)$, and the endogenous variables X_t influence Y_t in a linear fashion. We explicitly consider continuous and discrete variables in $X_t = (X_t^c, X_t^d)$, where $X_t^{c\prime} \in \Re^{K_c}$ are the continuous variables, $X_t^{d\prime} \in \Re^{K_d}$ discrete variables, and $K_c + K_d = K$. Similarly, $Z_{1t} = (Z_{1t}^c, Z_{1t}^d)$, $Z_{2t} = (Z_{2t}^c, Z_{2t}^d)$, where ${Z_{1t}^c}' \in \Re^{l_{1c}}$, ${Z_{2t}^c}' \in \Re^{l_{2c}}$ are the continuous variables and ${Z_{1t}^d}' \in \Re^{l_{1d}}$ and ${Z_{2t}^d}' \in \Re^{l_{2d}}$ discrete variables.

The partially linear model in (1) differs from the earlier semiparametric literature (Robinson (1988), Speckman (1988), Delgado and Mora (1995), and Härdle, Liang and Gao (2000)) in that it contains endogenous variables. It provides much needed flexibility through the nonparametric

[2] The 2SLS addresses the right hand side endogenous variables parametrically. Many recent papers consider nonparametric generalizations of 2SLS estimation to account for the endogeneity (see Matzkin (1994, 2008), Blundell and Powell (2003), Imbens and Newey (2009), and Dorolles et al. (2011)). The generalizations relax the tight parametric assumptions on the functional forms of structural equations, and thus estimation and inference are robust to potential model misspecifications. Though the flexibility is desirable in many applications, further assumptions on the structural equation or the nature of the endogeneity are needed to identify the parameters of interest. Furthermore, the pure nonparametric approach is generally associated with the "curse of dimensionality" to a large degree. On the other hand, our equation (1) offers a convenient alternative by modeling both nonparametric and parametric components, where the former provides robustness against misspecification, and the latter reduces the severeness of the "curse of dimensionality."

control function $m(Z_{1t})$, while the endogenous variables X_t enter the model parametrically, allowing for easy interpretation, implementation, and faster convergence. Consider again the estimation of the return to schooling in Card (1995). He estimates the structural form of an earnings model, where Y_t is the log of wages, X_t is the endogenous years of schooling, and Z_{1t} includes experience (potential experience, treated as exogenous) and indicators for race and residence. β can be interpreted as the return to schooling. Assuming $m(\cdot)$ to be a parametric function and using the proximity to a four-year college as IV for education[3], Card concludes with 2SLS estimation that OLS estimate for β might underestimate the return to schooling. However, we argue that the return to schooling parameter β might still be estimated inconsistently with 2SLS if $m(Z_{1t})$ is misspecified. Furthermore, assuming the linear reduced form in 2SLS creates another potential source of misspecification and can result in loss of efficiency (see equation (3) and the discussion). This calls for an efficient semiparametric instrumental variable estimation of model (1).

It is well known that the efficiency of the estimators for β becomes a concern when we relax the restrictive functional form assumptions on $m(Z_{1t})$ and on the reduced form $E(X_t|\mathbf{Z_t})$, for $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$. Newey (1990, 1993) considers efficient estimation of a parametrically specified structural model (m($\cdot$) is not present) with conditional moment restrictions using nearest neighbor and series estimators. Recently, many interesting papers consider efficient estimation of a semiparametric model more general than (1) using non-kernel based methods. Ai and Chen (2003) propose a semiparametric efficient estimator by the methods of minimum distance and sieves, which can be particularly convenient when $m(\cdot)$ enters the conditional moment expression in a nonlinear fashion, or when certain restrictions, such as additivity, are imposed on $m(\cdot)$. An empirical likelihood estimator has been considered by Otsu (2011), a penalized sieve minimum distance estimator has been proposed by Chen and Pouzo (2012), and a function space Tikhonov regularized minimum distance method has been studied by Florens et al. (2011) (see Chen and Pouzo (2012) for an excellent summary). However, implementing these efficient estimation methods usually entails a demanding numerical

[3]When instrumental variables are not available, one could use heteroskedasticity covariance restriction to identify and estimate the model as in Lewbel (2012).

optimization procedure[4]. Computationally intensive implementation may not lead to desirable finite sample performance.

The most commonly used kind of nonparametric regression estimation in econometrics is the kernel-based estimator, such as the Nadaraya-Watson, or the local polynomial one (Li and Racine (2007)). Kernel-based estimators provide attractive features of easy implementation with explicit algebraic structure and convenience in an asymptotic analysis. Li and Stengos (1996) and Baltagi and Li (2002) consider the efficient instrumental variable estimation of a semiparametric dynamic panel data model, but their estimators may not achieve the semiparametric efficiency bound. As far as we know, a kernel-based estimator of β with endogenous X_t in model (1) which reaches the semiparametric efficiency bound has not been formally considered.

In this paper we make two contributions to the literatures. First, we propose three new kernel-based and easy-to-implement estimators for β in model (1). We depart from the existing kernel-based estimation literature by modeling the reduced form nonparametrically. The estimators are consistent, asymptotically normal and reach the semiparametric efficiency bound. Exhibiting good finite sample performances, they provide a viable alternative that complements other estimators in the literature. Second, on the empirical side, we apply the proposed methods to estimate the return to schooling using data in Card (1995). The results help to resolve the puzzle on the estimates of the return to schooling.

Utilizing the partially linear nature in model (1), we construct efficient estimators directly with kernel-based methods[5]. We explicitly allow the endogenous and exogenous variables in equation (1) to be discrete or continuous, which facilitate its application to empirical research. The structural functions characterized by β and $m(\cdot)$ are easily identified (see assumption A1 and the discussion). The estimators are easily implemented without entailing numerical optimizations and have explicit algebraic structures.

[4]The general estimator in Ai and Chen involves numerical optimizations. One could use the profile sieve minimum distance procedure outlined in Blundell et al. (2007) as a computationally simpler alternative.

[5]One could construct an efficient minimum distance estimator using kernel-based method to approximate the unknown items locally, which could still involve numerical optimization procedures.

The explicit algebraic structure of our efficient estimators reveals that allowing nonlinearity in the reduced form $E(X_t|\mathbf{Z_t})$ could be an important step in semiparametric efficient estimation of β. First, the identification of β is achieved with the existence of a positive definite matrix $E(W_t'W_t)$ in assumption A1, where $W_t = E(X_t|\mathbf{Z_t}) - E(X_t|Z_{1t})$ depends on the reduced form (see also our assumption A6, and assumption 4.1 in Ai and Chen). Second, the semiparametric efficiency bound of β (see Chamberlain (1992), or the definition of J_0 in section 3) depends on $E(X_t|\mathbf{Z_t})$. Third, $E(X_t|\mathbf{Z_t})$ in the reduced form is generally of unknown form, since the structural form of Y_t and X_t could contain nonlinearity in the endogenous and/or exogenous variables of unknown form, or even if the form of nonlinearity is known but information on X_t's conditional distribution given $\mathbf{Z_t}$ is insufficient to parameterize $E(X_t|\mathbf{Z_t})$. Our simulation and empirical results also indicate that failing to allow the reduced form to be nonparametric can lead to misleading parameter and standard error estimates.

When exogenous X_t enters model (1), Li and Racine (2007, p237) note that the challenge for an efficient semiparametric estimator of β is the "curse of dimensionality." It requires estimation of a nonparametric model with dimension K and l_1 (the dimension of (X_t, Z_{1t})), while a consistent but not necessarily efficient estimation of β and $m(\cdot)$ involves only nonparametric estimation with dimension l_1. Therefore, the "curse of dimensionality" may prevent researchers from applying *efficient* estimation procedures to a partially linear model. When endogenous variables X_t enters model (1), we contend that the challenge of efficient estimation of β resides in estimating a nonparametric model whose dimension is determined by $l_1 + l_2$, the number of exogenous variables $\mathbf{Z_t}$, where l_2 is the number of excluded exogenous variables which can be used as additional IV's for X_t. Specifically, we address the challenge of estimating $m(\cdot)$, the heteroskedasticity function, and the reduced form nonparametrically by placing proper smoothness conditions on them (see assumptions A2 and A6). Utilizing the smoothing parameter (bandwidth) to converge to zero at proper speeds with increasing sample sizes, together with higher order kernel functions (see assumptions A3 and A5), we manage to control the errors generated from estimating the nonparametric functions, such that the estimators are $\sqrt{n}$ consistent and asymptotically normal. The degree of "curse of dimensionality" in our kernel-based estimation is determined only by the number of continuous variables in $\mathbf{Z_t}$, since discrete variables in $\mathbf{Z_t}$ do not slow down the convergence, which can be useful in empirical applications with many dummy variables in the exogenous variables. The first two estimators are efficient relative to those considered previously under conditional homoskedasticity and the last estimator is efficient under heteroskedasticity. They are asymptotically equivalent to semiparametric IV estimators that optimally select instrument variables, and are thus efficient in a class of semiparametric IV estimators with conditional moment restrictions. We further show that they reach the semiparametric efficiency bound. A Monte Carlo study illustrates that our estimators perform well relative to other estimators in finite samples. Thus, we conclude that our efficient kernel-based semiparametric IV estimators, with the ease of implementation, provide a viable alternative that complements the estimators available in the literature.

On the empirical side, we apply our estimators to estimate the return to schooling (β) using the data in Card (1995). The semiparametric IV estimates for β are much smaller than the 2SLS estimate, and are much more precise. The empirical evidence further suggests that the difference between our estimates and the 2SLS estimate largely arises from the misspecification in the parametric reduced form. These results provide a good explanation for the puzzle on the larger 2SLS estimate than the OLS counterparts on the return to schooling in the literature.

In what follows, we provide a detailed description of our semiparametric model and propose three estimators in Section 2, provide the asymptotic properties in Section 3, perform a Monte Carlo study to investigate the finite sample performance of the estimators and to compare with other alternatives in Section 4, estimate the return to schooling in Section 5, and conclude in Section 6. All tables and graphs are relegated to an appendix, and proofs are referred to in another appendix (for the review purpose only) and are also available in our working paper (Yao and Zhang (2012)).

2 Semiparametric Model

Consider the model in Equation (1) and assume the existence of instrumental variables $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$ with $E(\epsilon_t|\mathbf{Z_t}) = 0$, for all t. To motivate the estimation, suppose we know the true conditional expectation $E(Y_t|Z_{1t}) = m(Z_{1t}) + E(X_t|Z_{1t})\beta$. Hence, we could subtract $E(Y_t|Z_{1t})$ from (1) to obtain

$$Y_t - E(Y_t|Z_{1t}) = (X_t - E(X_t|Z_{1t}))\beta + \epsilon_t. \tag{2}$$

The conditional expectations are generally unknown, but we could replace them with nonparametric conditional mean estimators $\hat{E}(Y_t|Z_{1t})$ and $\hat{E}(X_t|Z_{1t})$. However, due to the correlation between ϵ_t and X_t, we can not apply Robinson (1988)'s estimator by regressing $Y_t - \hat{E}(Y_t|Z_{1t})$ on $X_t - \hat{E}(X_t|Z_{1t})$.

Following Davidson and MacKinnon (2004) and adapting to the current notation, we could estimate β by

$$\check{\beta}^{[1]} = (Q'\check{X})^{-1}Q'\check{Y}.$$

using the instrument variables $Q = \{Q_1', Q_2', \cdots, Q_n'\}'$ if $Q_t' \in \Re^K$ and Q_t contain Z_{1t}, Z_{2t}. Here

$$\check{X} = \begin{bmatrix} X_1 - \hat{E}(X_1|Z_{11}) \\ \vdots \\ X_n - \hat{E}(X_n|Z_{1n}) \end{bmatrix}, \check{Y} = \begin{bmatrix} Y_1 - \hat{E}(Y_1|Z_{11}) \\ \vdots \\ Y_n - \hat{E}(Y_n|Z_{1n}) \end{bmatrix}.$$

Li and Stengos' (1996) estimator sets $Q_t = Z_{2t} - \hat{E}(Z_{2t}|Z_{1t})$ in handling the endogeneity in the partially linear panel data models. In the case that $Q_t' \in \Re^{K+q}$ for some positive integer q, inspired by (8.29) in Davidson and MacKinnon, we consider the estimator of the form

$$\check{\beta}^{[2]} = (\check{X}'Q(Q'Q)^{-1}Q'\check{X})^{-1}\check{X}'Q(Q'Q)^{-1}Q'\check{Y}.$$

Estimators considered in Baltagi and Li (2002) are technically similar to $\check{\beta}$ using $Q_t = Z_{2t} - \hat{E}(Z_{2t}|Z_{1t})$, $Q_t = Z_{2t}$ or some other variables in the subspace $s(\vec{Z})$ spanned by the columns of $\vec{Z} = (\mathbf{Z_1}', \mathbf{Z_2}', \cdots, \mathbf{Z_n}')'$. As noted in Baltagi and Li (2002), setting $Q_t = Z_{2t} - \hat{E}(Z_{2t}|Z_{1t})$ is equivalent to using $Q_t = Z_{2t}$ because $E(Z_{2t}|Z_{1t})$ is orthogonal to $X_t - E(X_t|Z_{1t})$. The above cited papers investigate the partially linear panel data or dynamic panel data model and they essentially consider variables in $s(\vec{Z})$ as instrumental variables in Q. More generally, we could utilize $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$ as instrumental variables because they are by definition exogenous. Hence, we consider

$$\breve{\beta} = (\check{X}'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{X})^{-1}\check{X}'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{Y}. \tag{3}$$

In (3), $\check{X}$ is projected onto the subspace $s(\vec{Z})$ through the projection operator $\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'$, and $\mathbf{Z_t}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{X}$ estimates the conditional expectation $E(X_t - E(X_t|Z_{1t})|\mathbf{Z_t})$ (in Li and Stengos' estimator, one could interpret $\check{X}$ being projected parametrically onto the subspace spanned by $\{Z_{2t} - E(Z_{2t}|Z_{1t})\}_{t=1}^{n}$.). If the conditional expectation of $\check{X}$ given $\vec{Z}$ or the reduced form is not linear in $\vec{Z}$, and since the conditional mean is the optimal predictor of $\check{X}$ given $\vec{Z}$ in the mean square sense, we expect gains in efficiency by replacing the projection $\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{X}$ with a nonparametric estimate.

As indicated in Robinson (1988, p. 945), a valid instrument for $X_t - E(X_t|Z_{1t})$ is a vector of functions of $\mathbf{Z_t}$ that includes Z_{1t} and is independent of ϵ_t, such that the covariance matrix in the limiting distribution of the $n^{\frac{1}{2}}$ consistent estimator of β exists. One candidate for the instrument is $\hat{E}(X_t|\mathbf{Z_t}) - \hat{E}(X_t|Z_{1t})$, where $\hat{E}(X_t|\mathbf{Z_t})$ is a nonparametric estimator of $E(X_t|\mathbf{Z_t})$. We note that it can estimate $E(X_t - E(X_t|Z_{1t})|\mathbf{Z_t})$ consistently even if $E(X_t - E(X_t|Z_{1t})|\mathbf{Z_t})$ is not linear in $\mathbf{Z_t}$. When $m(Z_{1t})$ is absent but X_t has a nonparametric reduced form, this estimator is similar to those in Newey (1990, 1993) for nonlinear equations with known structural form but unknown reduced form. This approach has not been pursued formally in the literature.

The conditional expectation is generally of unknown form if the structural form of Y_t and X_t contains nonlinearities in the endogenous and/or exogenous variables of unknown form, or even if the form of nonlinearity is known but information on X_t's conditional distribution given $\mathbf{Z_t}$ is insufficient to parameterize $E(X_t|\mathbf{Z_t})$. Motivated by this observation, we propose the first estimator of β as

$$\hat{\beta} = (\hat{W}'\hat{W})^{-1}\hat{W}'(Y - \hat{E}(Y|\vec{Z}_1)). \tag{4}$$

Define the density of Z_{1t} at z_{10} as $f_1(z_{10})$, and the density of $\mathbf{Z_t}$ at z_0 as $f(z_0)$. We estimate them with the Rosenblatt density estimators with both continuous and discrete variables, and use the Nadaraya-Watson estimators for $E(A_0|z_{10})$, and $E(A_0|z_0)$. Specifically, they are

$$\hat{f}_1(z_{10}) = \frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\Big(\frac{Z_{1t}^c - z_{10}^c}{h_1}\Big) I(Z_{1t}^d = z_{10}^d),$$

$$\hat{E}(A_0|z_{10}) = \frac{\frac{1}{nh_1^{l_{1c}}} \sum_{t=1}^{n} K_1\big(\frac{Z_{1t}^c - z_{10}^c}{h_1}\big) I(Z_{1t}^d = z_{10}^d) A_t}{\hat{f}_1(z_{10})},$$

$$\hat{f}(z_0) = \frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\Big(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\Big) I(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d),$$

$$\hat{E}(A_0|z_0) = \frac{\frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{t=1}^{n} K_2\big(\frac{Z_{1t}^c - z_{10}^c}{h_2}, \frac{Z_{2t}^c - z_{20}^c}{h_2}\big) I(Z_{1t}^d = z_{10}^d, Z_{2t}^d = z_{20}^d) A_t}{\hat{f}(z_0)},$$

where h_1 and h_2 are bandwidths which go to zero as $n \to \infty$. $K_1(\cdot), K_2(\cdot)$, and $I(\cdot)$ are the kernel and indicator functions. Let

$$\hat{W} = \begin{bmatrix} \hat{W}_1 \\ \hat{W}_2 \\ \vdots \\ \hat{W}_n \end{bmatrix} = \begin{bmatrix} \hat{W}_{1,1} & \hat{W}_{1,2} & \vdots & \hat{W}_{1,K} \\ \hat{W}_{2,1} & \hat{W}_{2,2} & \cdots & \hat{W}_{2,K} \\ \cdots & \vdots & \vdots & \vdots \\ \hat{W}_{n,1} & \hat{W}_{n,2} & \cdots & \hat{W}_{n,K} \end{bmatrix}, \; \hat{W}_{t,k} = \hat{E}(X_{t,k}|\mathbf{Z_t}) - \hat{E}(X_{t,k}|Z_{1t}),$$

where $X_{t,k}$ is the kth element of random vector X_t. Define $Y = (Y_1, \cdots, Y_n)'$, $\vec{Z}_1 = (Z'_{11}, \cdots, Z'_{1n})'$ and $\hat{E}(Y|\vec{Z}_1) = (\hat{E}(Y_1|Z_{11}), \hat{E}(Y_2|Z_{12}), \cdots, \hat{E}(Y_n|Z_{1n}))'$.

In constructing $\hat{\beta}$, we first replace the unknown in $X_t - E(X_t|Z_{1t})$ by the Nadaraya-Watson estimator $\hat{E}(X_t|Z_{1t})$. Then we further estimate the conditional expectation $E(X_t - \hat{E}(X_t|Z_{1t})|\mathbf{Z_t})$. Geometrically, we project $X_t - \hat{E}(X_t|Z_{1t})$ onto $M(\mathbf{Z_t})$, the closed linear subspace of L^2 consisting of measurable function of $\mathbf{Z_t}$ with finite second moments. We only replace X_t with its conditional expectation estimator $\hat{E}(X_t|\mathbf{Z_t})$ because $E(X_t|Z_{1t})$ is already in $M(\mathbf{Z_t})$. Thus, this is similar to Theil's (1953) version of two stage least square estimator for simultaneous equations.

The second estimator we consider is

$$\tilde{\beta} = (\hat{W}'\hat{W})^{-1}\hat{W}'(\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1)), \tag{5}$$

where $\hat{E}(Y|\vec{Z}) = (\hat{E}(Y_1|\mathbf{Z_1}), \hat{E}(Y_2|\mathbf{Z_2}), \cdots, \hat{E}(Y_n|\mathbf{Z_n}))'$. In essence, relative to the first estimator, we further project Y_t onto $M(\mathbf{Z_t})$ nonparametrically. Hence, it is in the spirit of the traditional two stage least square estimator by Basmann (1957).

The first two estimators are shown in the next section to be consistent and asymptotically normally distributed (see Theorem 1 and the discussion). They are efficient relative to the estimator previously considered as demonstrated in Theorem 2 under conditional homoskedasticity. However, they do not exploit the structure of heteroskedasticity if it is present. To properly account for the information provided by heteroskedasticity, it is important to estimate the conditional variance correctly. We provide our third estimator $\tilde{\beta}^H$ in two steps, whose asymptotic properties are provided in Theorem 3. To simplify the analysis, we focus on the case that heteroskedasticity depends only on the included exogenous variables, that is, $E(\epsilon_t^2|\mathbf{Z_t}) = \sigma^2(Z_{1t})$ as in assumption A6(1) below. First, the estimated residual based on $\tilde{\beta}$ is $\tilde{\epsilon}_t = Y_t - \hat{E}(Y_t|Z_{1t}) - (X_t - \hat{E}(X_t|Z_{1t}))\tilde{\beta}$. The conditional variance is nonparametrically estimated as $\hat{\sigma}^2(Z_{1t}) = \hat{E}(\tilde{\epsilon}^2|Z_{1t})$. The conditional covariance matrix $\Omega(\vec{Z}_1)$ is estimated with $\hat{\Omega}(\vec{Z}_1)$, which is a diagonal matrix with the $t-th$ element as $\hat{\sigma}^2(Z_{1t})$. Second, inspired by the generalized least squares estimator, we construct the feasible efficient estimator as

$$\tilde{\beta}^H = (\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)(\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1)). \tag{6}$$

Using an indicator function to account for the discrete variables has been considered in other contexts by Delgado and Mora (1995), Fan et al. (1998), and Camlong-Viot et al. (2006). One could also follow Racine and Li (2004) to introduce a more delicate estimator for the discrete variables for improved finite sample performance, but in this case, one needs to consider the selection of additional smoothing parameters.

3 Asymptotic properties

Similar to the parametric instrumental variable estimation, our semiparametric instrumental variable estimators are likely to be biased. We investigate their asymptotic properties with the following assumptions.

A1: (1)$\{Y_t, X_t, \mathbf{Z_t}\}_{t=1}^n$ is an independent and identically distributed (iid) sequence of random vectors related as in Equation (1). (2) $E(\epsilon_t|\mathbf{Z_t}) = 0$ for all t.
(3) Let $W_t = E(X_t|\mathbf{Z_t}) - E(X_t|Z_{1t})$, $E(W_t'W_t)$ is a symmetric and positive definite matrix.

In A1(2), we require the conditional expectation of the error term ϵ_t given $\mathbf{Z_t}$ to be zero, but we allow ϵ_t's and X_t's to be possibly correlated, and thus $\mathbf{Z_t}$ plays the role of instrumental variables.

A1(3) is the identification assumption for β, similar to the identification assumption in Robinson (1988). From Equation (1) and assumption A1(2),

$$E(Y_t|\mathbf{Z_t}) - E(Y_t|Z_{1t}) = [E(X_t|\mathbf{Z_t}) - E(X_t|Z_{1t})]\beta = W_t\beta.$$

We pre-multiply both sides by W_t', take expectation, and obtain with assumption A1(3) that $\beta = (E(W_t'W_t))^{-1}E(W_t'(E(Y_t|\mathbf{Z_t}) - E(Y_t|Z_{1t})))$. Since conditional expectations are identified, β is identified with A1(3). We provide some intuitive implications. First, X_t cannot contain a constant. Second, it implies that $E(X_t|\mathbf{Z_t}) \neq E(X_t|Z_{1t})$. Since $\mathbf{Z_t} = (Z_{1t}, Z_{2t})$, any element of X_t cannot be perfectly a.s. predictable by Z_{1t}, i.e., X_t cannot be some function of Z_{1t} only. Obviously, Z_{2t} cannot simply be a linear combination of Z_{1t}, so Z_{2t} needs to contain variables that are linearly independent of Z_{1t}. A1(3) forbids more general forms of dependence. Third, because W_t cannot be a.s. zero, no elements of $E(X_t|\mathbf{Z_t}) - E(X_t|Z_{1t})$ are multicollinear. This fails if X_t is collinear.

We let C denote a generic constant below, which can vary from one place to another. Let $G = G_1^c \times G_1^d \times G_2^c \times G_2^d \subset \Re^l$, $G_1 = G_1^c \times G_1^d \subset \Re^{l_1}$ and $G_2 = G_2^c \times G_2^d \subset \Re^{l_2}$. G_1^c and G_2^c are compact, and G_1^d and G_2^d have finite support, i.e., they contain finite number of discrete elements.

Let's denote a generic function $g(Z_{1t}) \in C_1^s$ if $g(Z_{1t})$ is s times continuously differentiable w.r.t. Z_{1t}^c, with its sth order derivative uniformly continuous on G_1^c, and for $|j| = 1, 2, \cdots, s$, $\sup_{Z_{1t}\in G_1} |\frac{\partial^j}{\partial(Z_{1t}^c)^j} g(Z_{1t})| < \infty$. Here, the $|j|th$ order derivative is

$$\frac{\partial^j}{\partial(Z_{1t}^c)^j} g(Z_{1t}) \equiv \frac{\partial^{|j|} g(Z_{1t}^c, Z_{1t}^d)}{\partial(Z_{1t,1}^c)^{j_1}\partial(Z_{1t,2}^c)^{j_2}\cdots\partial(Z_{1t,l_{1c}}^c)^{j_{l_{1c}}}}.$$

We adopt the notation that $j = (j_1, j_2, \cdots, j_{l_{1c}})'$, $|j| = \sum_{i=1}^{l_{1c}} j_i$. In what follows, we denote $\sum_{0\le|j|\le s} = \sum_{|j|=0} + \sum_{|j|=1} + \cdots + \sum_{|j|=s}$, $j! = j_1! \times j_2! \times \cdots \times j_{l_{1c}}!$, $(Z_{1t}^c)^j = (Z_{1t,1}^c)^{j_1} \times (Z_{1t,2}^c)^{j_2} \times \cdots \times (Z_{1t,l_{1c}}^c)^{j_{l_{1c}}}$, where $Z_{1t,i}^c$ refers to the $i-th$ element in Z_{1t}^c.

Denote a generic function $g(Z_{1t}) \in C_1^{s_1}$ if $g(Z_{1t})$ is s_1 times continuously differentiable w.r.t. Z_{1t}^c, with its sth and s_1th order derivative uniformly continuous on G_1^c, and for $|j| = 1, 2, \cdots, s_1$, $\sup_{Z_{1t}\in G_1} |\frac{\partial^j}{\partial(Z_{1t}^c)^j} g(Z_{1t})| < \infty$. Denote a generic function $g(\mathbf{Z_t}) \in C_{1,2}^{s_1}$ if $g(\mathbf{Z_t})$ is s_1 times continuously differentiable w.r.t. $\mathbf{Z_t^c} = (Z_{1t}^c, Z_{2t}^c)$, with its sth and s_1th order derivative uniformly continuous on $G^c = G_1^c \times G_2^c$, and for $|j| = 1, 2, \cdots, s_1$, $\sup_{\mathbf{Z_t} \in G} |\frac{\partial^j}{\partial(\mathbf{Z_t^c})^j} g(\mathbf{Z_t})| < \infty$. Here, we denote the $|j|th$ order derivative as

$$\frac{\partial^j}{\partial(\mathbf{Z_t^c})^j} g(\mathbf{Z_t}) \equiv \frac{\partial^{|j|} g(Z_{1t}^c, Z_{1t}^d, Z_{2t}^c, Z_{2t}^d)}{\partial(Z_{1t,1}^c)^{j_1} \partial(Z_{1t,2}^c)^{j_2} \cdots \partial(Z_{1t,l_{1c}}^c)^{j_{l_{1c}}} \partial(Z_{2t,1}^c)^{j_{l_{1c}+1}} \partial(Z_{2t,2}^c)^{j_{l_{1c}+2}} \cdots \partial(Z_{2t,l_{2c}}^c)^{j_{l_{1c}+l_{2c}}}}.$$

A2: (1) Denote the density of Z_{1t} at z_{10} by $f_1(z_{10}^c, z_{10}^d)$. ($f_1(z_{10}^c, z_{10}^d)$ is the "mixed joint density", defined with respect to the product measure on the respective support of z_{10}^c and z_{10}^d. One can construct it as the product of conditional density of z_{10}^c given z_{10}^d and the marginal probability function of z_{10}^d.) Assume $f_1(z_{10}^c, z_{10}^d) \in C_1^s$ $\forall z_{10}^d \in G_1^d$. (2) $0 < C < f_1(z_1^c, z_1^d) < \infty$, for all $z_1 \in G_1$.
(3) $X_{t,k} = E(X_{t,k}|Z_{1t}) + X_{t,k} - E(X_{t,k}|Z_{1t}) = g_{1,k}(Z_{1t}) + e_{1,kt}$. $\forall z_{10}^d \in G_1^d$, $g_{1,k}(z_{10}) \in C_1^{s_1}$. The conditional density of Z_{1t} given $e_{1,kt}$ is bounded, and the conditional density of $X_{t,k}$ given Z_{1t} is continuous around Z_{1t}^c.
(4) Denote the density of $\mathbf{Z_t}$ at $\mathbf{z_0}$ by $f_z(\mathbf{z_0})$. $\forall z_{10}^d \in G_1^d$, $z_{20}^d \in G_2^d$, $f_z(\mathbf{z_0}) \in C_{1,2}^{s_1}$.
(5) $0 < C < f_z(\mathbf{z_0}) < \infty$, for all $\mathbf{z_0} \in G$.
(6) $X_{t,k} = E(X_{t,k}|\mathbf{Z_t}) + X_{t,k} - E(X_{t,k}|\mathbf{Z_t}) = g_k(\mathbf{Z_t}) + e_{kt}$. $\forall z_{10}^d \in G_1^d$, $z_{20}^d \in G_2^d$, $g_k(\mathbf{z_0}) \in C_{1,2}^{s_1}$. The conditional density of $\mathbf{Z_t}$ given e_{kt} is bounded, and the conditional density of $X_{t,k}$ given $\mathbf{Z_t}$ is continuous around $\mathbf{Z_t^c}$. (7) $m(z_{10}) \in C_1^{s_1}$.

A3: (1) For $x \in \Re^d$, $d = l_{1c}$ or l_{2c}, the kernel function $K(x)(K_1(x)$ or $K_2(x))$ is bounded with bounded support, and it is of order $3s_1$. (2) $|u^i K(u) - v^i K(v)| \le C_K ||u - v||$, $|i| = 0, 1, 2, \cdots, s_1$.

A4: (1) For some $\delta > 0$, $E(|X_{t,k}|^{2+\delta}|\mathbf{Z_t}), E(|X_{t,k}|^{2+\delta}|Z_{1t}) < \infty$, $|g_k(\mathbf{Z_t})|, |g_{1,k}(Z_{1t})| < \infty$ almost everywhere. (2) $E(|\epsilon_t|^{2+\delta}|\mathbf{Z_t}), E(|\epsilon_t|^{2+\delta}|Z_{1t}) < \infty$. (3) $E(\epsilon_t^2|\mathbf{Z_t}) = \sigma^2(\mathbf{Z_t})$.
(4) The conditional density of Z_{1t} given ϵ_t is bounded, and the conditional density of ϵ_t given Z_{1t} is continuous at Z_{1t}^c. The conditional density of $\mathbf{Z_t}$ given ϵ_t is bounded, and the conditional density of ϵ_t given $\mathbf{Z_t}$ is continuous at $\mathbf{Z_t^c}$.

A5: as $n \to \infty$, (1) $nh_1^{2l_{1c}} \to \infty$. (2) $nh_2^{2(l_{1c}+l_{2c})} \to \infty$. (3) $nh_1^{2(s+1)}, nh_2^{2(s_1+1)} \to 0$.

We require that the densities $f_1(z_{10})$ and $f(z_0)$ to be bounded away from zero on a bounded support set in assumption A2 to handle the technical difficulty due to the random denominators $\hat{f}(z_{10})$ and $\hat{f}(z_0)$. Alternatively, we could follow Robinson (1988), Bickel (1982), and Manski (1984) to "trim" out small $\hat{f}(z_{10})$ and $\hat{f}(z_0)$, or replace them with a small but positive constant. We expect that this will not change the asymptotic results so we will not introduce it in the definition. As another strategy, we could follow Li and Stengos (1996) or Powell et al. (1989) to estimate a density-weighted relationship. We expect that the asymptotic results will be different.

Assumptions A2(1), (2), (4), (5), and (7) require the densities $f_1(z_{10})$, $f_z(\mathbf{z_0})$, and $m(z_{10})$ to be bounded and continuously differentiable w.r.t. its continuous components. These assumptions are commonly used in nonparametric kernel regression, enabling the use of Taylor expansion. They are similar in spirit to the smoothness and boundedness condition in Definition 2 of Robinson (1988), or the assumption A1 of Li and Stengos (1996). A2(3) and (6) explicitly assume the relationship between X_t and Z_{1t} and between X_t and $\mathbf{Z_t}$. Similar assumptions have been maintained in Speckman (1988) in the fixed design case. Assumption A3 requires the kernel function to be smooth and bounded (Martins-Filho and Yao (2007)). Asymptotic distributions are established using Liapunov's central limit theorem, with conditional moments assumption of ϵ_t and X_t given $\mathbf{Z_t}$ or Z_{1t} in A4. The bandwidth assumptions A5(1) and (2) are in line with those used in the literature (Martins-Filho and Yao (2007)). A3, together with A5(3), specifies the kernel properties and the rate of decay for the bandwidths. They are used to control the bias introduced in the nonparametric regression, which is similar to assumptions in, for example, Robinson (1988) and Li and Stengos (1996). However, A5(3) is stronger than that maintained in Li and Stengos, or Robinson. As our estimators involve estimation of W_t, the bias arises not only from estimation of $E(X_t|Z_{1t})$, but also from estimation of $E(X_t|\mathbf{Z_t})$. A5 requires choosing a higher order kernel to eliminate the bias asymptotically. A5 illustrates the extra technical assumption needed to perform *efficient* semiparametric endogenous variables estimation, relative to consistent semiparametric estimation indicated in the introduction. Results in Theorems 1-2 are obtained for general heteroskedasticity structure in A4(3).

A6: (1) $E(\epsilon_t^2|\mathbf{Z_t}) = E(\epsilon_t^2|Z_{1t}) = \sigma^2(Z_{1t})$. $0 < C < \sigma^2(Z_{1t}) < \infty$, $\sigma^2(z_{10}) \in C_1^{s_1}\ \forall z_{10}^d \in G_1^d$.
(2) $E(\frac{1}{\sigma^2(Z_{1t})} W_t' W_t)$ is a symmetric and positive definite matrix.
(3) $E(|X_{t,k}|^{4+\delta}|Z_{1t}) < \infty$, and $E(|\epsilon_t|^{4+\delta}|Z_{1t}) < \infty$. The conditional density of Z_{1t} given $|X_{t,k}\epsilon_t|$ is bounded, and the conditional density of $|X_{t,k}\epsilon_t|$ given Z_{1t} is continuous at Z_{1t}^c. The conditional density of Z_{1t} given $e_{1,kt}e_{1,k't}$ is bounded, and the conditional density of $|X_{t,k}X_{t,k'}|$ given Z_{1t} is continuous at Z_{1t}^c for all $k, k' \in \{1, \cdots, K\}$.

A7: (1) $E(X_{t,k}^4|\mathbf{Z_t}) < \infty$. (2) $E(\epsilon_t^4|\mathbf{Z_t}) < \infty$.

For efficient estimation, we restrict the structure of $E(\epsilon_t^2|\mathbf{Z_t})$ to be $\sigma^2(Z_{1t})$ in A6(1) for simplicity, so the heteroskedasticity depends only on the included exogenous variables. Assumption A6 provides higher moments and additional smoothness conditions, enabling us to obtain the asymptotic results for $\tilde{\beta}^H$, which involves estimation of the conditional covariance matrix of ϵ_t. The asymptotic results for $\hat{\beta}$ are obtained with additional moments conditions in A7. Lemma 1 in the Appendix of our working paper (Yao and Zhang (2012)) establishes the order in probability of certain linear combinations of kernel functions that appear repeatedly in the component expressions of our estimators. We use it in the proofs of the Theorems.

Theorem 1 *Let Φ_0 be a $K \times K$ positive definite matrix with the (i,j)th element $E[\sigma^2(\mathbf{Z_t})(g_i(\mathbf{Z_t}) - E_{Z_{2\tau}|Z_{1t}}(g_i(Z_{1t}, Z_{2\tau})))(g_j(\mathbf{Z_t}) - E_{Z_{2\tau}|Z_{1t}}(g_j(Z_{1t}, Z_{2\tau})))]$, where $E_{Z_{2\tau}|Z_{1t}}(.)$ denotes the conditional expectation of $Z_{2\tau}$ given Z_{1t}. Given assumptions A1-A5, we have*

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} N(0, (E(W_t'W_t))^{-1}\Phi_0(E(W_t'W_t))^{-1}).$$

$\hat{\beta}$ is relatively easier to construct because it does not involve further projecting Y_t onto $M(\mathbf{Z_t})$ and is in spirit similar to Theil's two stage least square estimator. However, we notice that $\hat{W}_t$ estimates W_t nonparametrically, so the simplicity in Theil's original estimator disappears. In finite samples, $\hat{\beta}$ and $\tilde{\beta}$ are not unbiased. Furthermore, compared to $\tilde{\beta}$, we find that the asymptotic expansion of $\hat{\beta} - \beta$ involves more stochastic terms, whose magnitudes need to be further controlled with the additional assumption A7 to obtain the asymptotic distribution. However, we do find the asymptotic distribution of $\hat{\beta}$ to be the same as that of $\tilde{\beta}$. The finite sample performance of the two proposed estimators is investigated in the simulation section and it indicates that $\tilde{\beta}$ always outperforms $\hat{\beta}$. For this reason, we do not provide asymptotic result for $\hat{\beta}$ formally to save space.

From Theorem 1 and discussion above, we note $\tilde{\beta}$ and $\hat{\beta}$ are consistent and asymptotically normal with the general conditional variance structure specified in A4(3). When ϵ_t is conditionally homoskedastic, i.e., $V(\epsilon_t|\mathbf{Z_t}) = \sigma_0^2$, it is straightforward to compare the asymptotic properties of the proposed estimators with $\check{\beta}$ because they are all consistent and converge to a normal distribution at rate $\sqrt{n}$. We find the asymptotic variance of $\check{\beta}$ is always greater or equal to that of the two proposed estimators, i.e., the difference is a positive semidefinite matrix. Thus, $\tilde{\beta}$ and $\hat{\beta}$ are asymptotically efficient relative to $\check{\beta}$.

Theorem 2 *If $V(\epsilon_t|\mathbf{Z_t}) = \sigma_0^2$, then the asymptotic variance of $\check{\beta}$ is greater than or equal to that of $\hat{\beta}$ or $\tilde{\beta}$.*

To illustrate the point, we first observe that in Φ_0,

$$E_{Z_{2\tau}|Z_{1t}}(g_j(Z_{1t}, Z_{2\tau})) = E_{Z_{2\tau}|Z_{1t}}(E(X_{t,j}|(Z_{1t}, Z_{2\tau}))) = E(X_{t,j}|Z_{1t}).$$

Therefore, $(E(W_t'W_t))^{-1}\Phi_0(E(W_t'W_t))^{-1} = \sigma_0^2(E(W_t'W_t))^{-1}$. For the estimator $\check{\beta}$ defined in Equation (3), suppose the estimated items are replaced by the true values, we obtain with Equation (2),

$$\begin{aligned} \check{\beta}^{Tr} &= ((\check{X}^{Tr})'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{X}^{Tr})^{-1}(\check{X}^{Tr})'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{Y}^{Tr} \\ &= \beta + ((\check{X}^{Tr})'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\check{X}^{Tr})^{-1}(\check{X}^{Tr})'\vec{Z}(\vec{Z}'\vec{Z})^{-1}\vec{Z}'\vec{\epsilon} \end{aligned}$$

where $\check{X}^{Tr} = ((X_1 - E(X_1|Z_{11}))', (X_2 - E(X_2|Z_{12}))', \cdots, (X_n - E(X_n|Z_{1n}))')'$, $\check{Y}^{Tr} = ((Y_1 - E(Y_1|Z_{11}))', \cdots, (Y_n - E(Y_n|Z_{1n}))')'$, and $\vec{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$. We expect the asymptotic distribution of $\check{\beta}$ to be the same as that of $\check{\beta}^{Tr}$, if the unknown conditional expectations are estimated with nonparametric kernel estimates. Suppose that $\frac{1}{n}\vec{Z}'\vec{\epsilon} \xrightarrow{p} 0$, which follows from the assumption of $\mathbf{Z_t}$ being instrumental variables. Assume further that $\frac{1}{n}((\check{X}^{Tr})'Z) \xrightarrow{p} E[(X_t - E(X_t|Z_{1t}))'\mathbf{Z_t}] = A$, and $\frac{1}{n}Z'Z \xrightarrow{p} E(\mathbf{Z_t}'\mathbf{Z_t}) = B$, A has rank K and B is positive definite. We have $\frac{1}{n}V(\vec{Z}'\vec{\epsilon}) = \sigma_0^2 B$, because ϵ_t is conditionally homoskedastic. Thus, we expect

$$\sqrt{n}(\check{\beta}^{Tr} - \beta) \xrightarrow{d} N(0, \sigma_0^2(AB^{-1}A')^{-1}).$$

The theorem is proved if $\sigma_0^2(AB^{-1}A')^{-1} - \sigma_0^2(E(W_t'W_t))^{-1}$ is positive semidefinite.

The above claim is equivalent to $E(W_t'W_t) - AB^{-1}A'$ being positive semidefinite. We note $A = E[(X_t - E(X_t|Z_{1t}))'\mathbf{Z_t}] = E[(E(X_t'|\mathbf{Z_t}) - E(X_t'|Z_{1t}))\mathbf{Z_t}] = E(W_t'\mathbf{Z_t})$. Hence,

$$\begin{aligned} & E(W_t'W_t) - AB^{-1}A' = E(W_t'W_t) - E(W_t'\mathbf{Z_t})(E\mathbf{Z_t}'\mathbf{Z_t})^{-1}E(\mathbf{Z_t}'W_t) \\ = \; & E[W_t'(W_t - \mathbf{Z_t}(E\mathbf{Z_t}'\mathbf{Z_t})^{-1}E(\mathbf{Z_t}'W_t))] \\ = \; & E[(W_t - \mathbf{Z_t}(E\mathbf{Z_t}'\mathbf{Z_t})^{-1}E(\mathbf{Z_t}'W_t))'(W_t - \mathbf{Z_t}(E\mathbf{Z_t}'\mathbf{Z_t})^{-1}E(\mathbf{Z_t}'W_t))] \end{aligned}$$

which is positive semidefinite. The last equality is true as

$$\begin{aligned} & E[(\mathbf{Z_t}(E(\mathbf{Z_t}'\mathbf{Z_t}))^{-1}E(\mathbf{Z_t}'W_t))'(W_t - \mathbf{Z_t}(E(\mathbf{Z_t}'\mathbf{Z_t}))^{-1}E(\mathbf{Z_t}'W_t))] \\ = & E(W_t'\mathbf{Z_t})(E(\mathbf{Z_t}'\mathbf{Z_t}))^{-1}E(\mathbf{Z_t}'W_t) - E(W_t'\mathbf{Z_t})(E(\mathbf{Z_t}'\mathbf{Z_t}))^{-1}E(\mathbf{Z_t}'\mathbf{Z_t})(E(\mathbf{Z_t}'\mathbf{Z_t}))^{-1}E(\mathbf{Z_t}'W_t) \\ = & 0. \end{aligned}$$

When ϵ_t is conditionally heteroskedastic as in A6(1), we consider the feasible estimator $\tilde{\beta}^H$, which is based on the first stage estimation with $\tilde{\beta}$. We did not consider $\hat{\beta}$ because it is outperformed by $\tilde{\beta}$ in the simulation study in Section 4. The asymptotic property of $\tilde{\beta}^H$ is provided in Theorem 3.

Theorem 3 *If we assume A1-A6, then*

$$\sqrt{n}(\tilde{\beta}^H - \beta) \xrightarrow{d} N(0, (E(\frac{1}{\sigma^2(Z_{1t})}W_t'W_t))^{-1}).$$

It is straightforward to show that under heteroskedasticity in A6(1), $(E(W_t'W_t))^{-1}\Phi_0(E(W_t'W_t))^{-1} - (E(\frac{1}{\sigma^2(Z_{1t})}W_t'W_t))^{-1}$ is a positive semidefinite matrix. Thus, asymptotically, $\tilde{\beta}^H$ is efficient relative to $\tilde{\beta}$ and $\hat{\beta}$.

To shed light on the theoretical results above, let us consider a class of semiparametric IV estimators based on the model in Equation (1) that satisfies the conditional moment restriction $E(\epsilon_t|\mathbf{Z_t}) = 0$, where $\epsilon_t = \epsilon_t(\beta) = Y_t - E(Y_t|Z_{1t}) - (X_t - E(X_t|Z_{1t}))\beta$ as in Equation (2). Then, $\tilde{\beta}$ and $\hat{\beta}$ are asymptotically equivalent to the semiparametric IV estimator that optimally selects instrument variables. Thus, $\tilde{\beta}$ and $\hat{\beta}$ are efficient among this class of semiparametric IV estimators in the sense that their asymptotic variance is smallest. To establish this, suppose $E(Y_t|Z_{1t})$ and $E(X_t|Z_{1t})$ are known, or can be consistently estimated at a certain rate, then let $H(\mathbf{Z_t})$ denote an $h \times 1$ vector of functions of $\mathbf{Z_t}$, $h \geq K$. By law of iterated expectation we have $E(H(\mathbf{Z_t})\epsilon_t(\beta)) = 0$. Following Newey (1993), we could construct the IV estimators using the method of moments estimator. It is defined as

$$\beta^{MIV} = argmin_\beta \hat{g}_n(\beta)'\hat{P}\hat{g}_n(\beta), \quad \hat{g}_n(\beta) = \frac{1}{n}\sum_{t=1}^{n} H(\mathbf{Z_t})\epsilon_t(\beta),$$

for $h \times h$ positive semidefinite matrix $\hat{P}$, which may be random. Since $\epsilon_t(\beta)$ is linear in β, by solving the minimization problem we easily obtain,

$$\begin{aligned} \sqrt{n}(\beta^{MIV} - \beta) = & [\tfrac{1}{n}\textstyle\sum_t (X_t - E(X_t|Z_{1t}))'H(\mathbf{Z_t})'\hat{P}\tfrac{1}{n}\sum_t H(\mathbf{Z_t})(X_t - E(X_t|Z_{1t}))]^{-1} \\ & \times \tfrac{1}{n}\textstyle\sum_t (X_t - E(X_t|Z_{1t}))'H(\mathbf{Z_t})'\hat{P}\sqrt{n}\tfrac{1}{n}\sum_t H(\mathbf{Z_t})\epsilon_t, \end{aligned}$$

Assume $\hat{P} \xrightarrow{p} P$, where P is a positive semi-definite matrix, $\frac{1}{n}\sum_t H(\mathbf{Z_t})(X_t - E(X_t|Z_{1t})) \xrightarrow{p} -E(H(\mathbf{Z_t})\frac{\partial \epsilon_t(\beta)}{\partial \beta}) = -G$, and define $V = EH(\mathbf{Z_t})\epsilon_t\epsilon_t' H(\mathbf{Z_t})' = E\sigma^2(\mathbf{Z_t})H(\mathbf{Z_t})H(\mathbf{Z_t})'$, then

$$\sqrt{n}(\beta^{MIV} - \beta) \xrightarrow{d} N(0, (G'PG)^{-1}G'PVP'G(G'PG)^{-1}).$$

Under conditional homoskedasticity, $\sigma^2(\mathbf{Z_t}) = \sigma_0^2$, so the asymptotic variance of $\tilde{\beta}$ is $\sigma_0^2(E(W_t'W_t))^{-1}$. Let $A = G'PH(\mathbf{Z_t})\epsilon_t(\beta)$, $B = -[E(X_t - E(X_t|Z_{1t})|\mathbf{Z_t})]'\frac{\epsilon_t(\beta)}{\sigma_0^2}$, then

$$\begin{aligned} & (G'PG)^{-1}G'PVP'G(G'PG)^{-1}) - \sigma_0^2(E(W_t'W_t))^{-1} \\ = \quad & E\{(EAB')^{-1}[A - (EAB')(EBB')^{-1}B][A' - B'(EBB')^{-1}(EBA')](EBA')^{-1}.\} \end{aligned}$$

which is a quadratic form, so the difference is positive semidefinite. We note the asymptotic variances will be the same if we let the optimal instrumental variable to be $H(\mathbf{Z_t}) = -\frac{W_t'}{\sigma_0^2}$. Thus, asymptotically, $\tilde{\beta}$ and $\hat{\beta}$ behave like an optimal semiparametric IV estimator. Under conditional heteroskedasticity specified in A6(1), we could follow a similar argument above to show the difference between $(G'PG)^{-1}G'PVP'G(G'PG)^{-1}$ and $(E(\frac{1}{\sigma^2(Z_{1t})}W_t'W_t))^{-1}$ is positive semidefinite. Here, the two asymptotic variances will be the same if we let $H(\mathbf{Z_t}) = -\frac{W_t'}{\sigma^2(Z_{1t})}$. Thus, asymptotically, $\tilde{\beta}^H$ behaves similar to an optimal semiparametric IV estimator under heteroskedasticity.

We further compare the theoretical results in Theorems 1-3 with the semiparametric efficiency bound derived in Chamberlain (1992). Our model in Equation (1) and assumption A1(2) consider the estimation of β in the model of $\epsilon_t = Y_t - m(Z_{1t}) - X_t\beta$ with the conditional moment restriction $E(\epsilon_t|\mathbf{Z_t}) = 0$. Since $\frac{\partial \epsilon_t}{\partial \beta'} = -X_t$, define $D_0(\mathbf{Z_t}) \equiv E(\frac{\partial \epsilon_t}{\partial \beta'}|\mathbf{Z_t}) = -E(X_t|\mathbf{Z_t})$, $\Sigma_0(\mathbf{Z_t}) \equiv E(\epsilon_t\epsilon_t'|\mathbf{Z_t}) = \sigma^2(\mathbf{Z_t})$ with the general heteroskedasticity structure in A4(3), $H_0(\mathbf{Z_t}) \equiv E(\frac{\partial \epsilon_t}{\partial r}|\mathbf{Z_t}) = -1$ for $r = m(Z_{1t})$. The Fisher information bound for β is

$$\begin{aligned} J_0 \quad &= E\{E(D_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}D_0(\mathbf{Z_t})|Z_{1t}) - E(D_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}H_0(\mathbf{Z_t})|Z_{1t}) \\ &\times [E(H_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}H_0(\mathbf{Z_t})|Z_{1t})]^{-1}E(H_0(\mathbf{Z_t})'\Sigma_0(\mathbf{Z_t})^{-1}D_0(\mathbf{Z_t})|Z_{1t})\}. \end{aligned}$$

It is easy to show that under homoskedasticity $\Sigma_0(\mathbf{Z_t}) = \sigma_0^2$,

$$J_0 = \frac{1}{\sigma_0^2}E(W_t'W_t) = [(E(W_t'W_t))^{-1}\Phi_0(E(W_t'W_t))^{-1}]^{-1}.$$

as in Theorems 1. Under the heteroskedasticity structure imposed in A6(1), $\Sigma_0(\mathbf{Z_t}) = \sigma^2(Z_{1t})$, we have $J_0 = E(\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)$, as in Theorem 3. We therefore conclude that the estimators $\tilde{\beta}, \hat{\beta}$, and

$\tilde{\beta}^H$ reach the semiparametric efficiency bound. Efficient estimation actually calls for nonparametric estimation of $E(X_t|\mathbf{Z_t})$ since it shows up in J_0.

4 Monte Carlo Study

In this section, we perform a Monte Carlo study to implement our efficient semiparametric instrumental variable estimators and illustrate their finite sample performance. For ease of comparison, we consider the data-generating process in Baltagi and Li (2002) adapted to the iid set-up as

$$Y_t = \beta X_t + \alpha_1 Z_{1t} + \alpha_2 Z_{1t}^2 + \epsilon_t, \text{ and } X_t = g_i(Z_{1t}, Z_{2t}) + U_t.$$

Here, the nonlinear function $m(Z_{1t})$ is $\alpha_1 Z_{1t} + \alpha_2 Z_{1t}^2$ and we fix $\beta = \alpha_1 = \alpha_2 = 1$. We generate Z_{1t} and Z_{2t} independently from a standard normal distribution, truncated to $[-1, 1]$. Conditional on $\mathbf{Z_t}$, ϵ_t and U_t are generated from a bivariate normal distribution with zero mean, variance $\sigma_i^2(\mathbf{Z_t})$, and correlation θ. We truncate ϵ_t and U_t to $[-1,1] \times [-1,1]$. We consider $\sigma_1^2(\mathbf{Z_t}) = 1$ for the homoskedasticity case, and $\sigma_2^2(\mathbf{Z_t}) = Z_{1t}^2$ for the heteroskedasticity case. We choose $\theta = 0.2, 0.5$, and 0.8. As θ increases, the correlation between X_t and ϵ_t increases, and thus endogeneity is magnified. We select two functions for g_i, with $g_1(Z_{1t}, Z_{2t}) = \gamma_1 Z_{1t} + \gamma_2 Z_{2t}$ and $g_2(Z_{1t}, Z_{2t}) = \gamma_1 Z_{1t}^2 + \gamma_2 Z_{2t}^2$, with $\gamma_1 = \gamma_2 = 1$. Hence, with g_1, X_t depends on $\mathbf{Z_t}$ linearly, while in g_2, X_t relates to $\mathbf{Z_t}$ in a nonlinear fashion. It is easy to verify that assumptions maintained in A1, A2, and A4 are satisfied. We consider two sample sizes, $n = 100$ and 200, and perform 1000 repetitions for each experimental design.

To implement our estimators $\hat{\beta}$, $\tilde{\beta}$, and $\tilde{\beta}^H$, we need to select the bandwidth sequences h_1 and h_2. We select the bandwidth $\hat{h}_1$ using the *rule-of-thumb* data driven plug-in method of Ruppert, Sheather, and Wand (1995). We select $\hat{h}_2$ using $1.25SD(\mathbf{Z_t})n^{-1/6}$, where $SD(\mathbf{Z_t})$ is the standard deviation of $\mathbf{Z_t}$. We choose a second order Epanechnikov kernel, which satisfies our assumption A3(2) and part of A3(1). Though our assumption calls for a higher order kernel, it is known that in finite sample applications, nonnegative second order kernels have often yielded more stable estimation results as higher order kernel could generate negative weights to data. Thus, we investigate the robustness of our estimators with a popular second kernel function. From our asymptotic analysis in Theorem 1 and the comments before Theorem 1, $\hat{\beta}$ and $\tilde{\beta}$ might have high finite sample bias.

Aside from our proposed estimators, we also include the semiparametric estimator $\breve{\beta}^{(1)}$ without considering the endogenous variable as in Robinson, estimator $\breve{\beta}^{(2)}$ as in Li and Stengos with instrumental variable $Q_t = Z_{2t} - \hat{E}(Z_{2t}|Z_{1t})$ using the density weighted estimation, and $\breve{\beta}$ in Equation (3). $\breve{\beta}^{(1)}$ serves as the benchmark because it ignores the endogeneity problem. We evaluate the performance of each estimator using bias (B), standard deviation (S), and root mean squared error (R) as criteria. The results of the experiments with $\sigma_1^2(\mathbf{Z_t}) = 1$ are summarized in Table 1 in the Appendix for g_1 and Table 2 for g_2. The results with $\sigma_2^2(\mathbf{Z_t}) = Z_{1t}^2$ are provided in Tables 3 and 4.

We observe that $\breve{\beta}$ and $\breve{\beta}^{(2)}$ generally have negative bias under g_1, while other estimators produce positive bias. As the sample size n increases, the estimators' performance generally improves in terms of smaller bias, standard deviation, and root mean squared error. This observation is consistent with our asymptotic results in Theorems 1 and 3 that $\hat{\beta}$, $\tilde{\beta}$, and $\tilde{\beta}^H$ are consistent. Exceptions occur with $\breve{\beta}^{(1)}$ whose bias does not seem to drop with a large sample. Exceptions also exist with $\breve{\beta}^{(2)}$ and $\breve{\beta}$ in the data-generating process g_2. This observation is consistent with the fact that in g_2, $E(X_t|\mathbf{Z_t})$ is not linear in $\mathbf{Z_t}$, while $\breve{\beta}^{(2)}$ and $\breve{\beta}$ implicitly assume $E(X_t|\mathbf{Z_t})$ is linear in $\mathbf{Z_t}$.

As θ increases, the endogeneity problem is magnified. The bias of $\breve{\beta}^{(1)}$ increases. Although its standard deviation drops slightly, the drop in standard deviation is dominated by the increase in bias and the root mean squared error increases. This is expected because $\breve{\beta}^{(1)}$ does not take endogeneity into consideration. We notice that the bias of all estimators generally increases with θ. For $\tilde{\beta}$ and $\tilde{\beta}^H$ in the homoskedasticity experiments, their standard deviation drops with θ, though their root mean squared error increases.

When g_1 is in the data-generating process and homoskedasticity is present, we note $E(X_t|\mathbf{Z_t})$ is linear in $\mathbf{Z_t}$. In terms of relative performance, $\breve{\beta}^{(1)}$ carries the largest bias, but smallest standard deviation. As expected, its root mean squared error is largest with $\theta = 0.5$ or 0.8. Among the other estimators that take endogeneity into account, $\breve{\beta}^{(2)}$'s or $\breve{\beta}$'s bias is smallest, followed by $\tilde{\beta}$, by $\tilde{\beta}^H$, and $\hat{\beta}$. The relatively large bias of $\tilde{\beta}$, $\tilde{\beta}^H$, and $\hat{\beta}$ is expected as we use a second order nonnegative kernel function, which is often found to yield more stable estimation results than their higher order counterparts. However, $\tilde{\beta}$ is the best using standard deviation as the criterion, followed by $\tilde{\beta}^H$ and $\hat{\beta}$, which is better for a large θ, or followed by $\check{\beta}$, while $\check{\beta}^{(2)}$ carries the largest standard deviation. This result is consistent with Theorems 1-3, which indicate the asymptotic variance of $\tilde{\beta}$, $\tilde{\beta}^H$, and $\hat{\beta}$ is smaller than or equal to $\check{\beta}$. Although having the same asymptotic distribution, $\hat{\beta}$'s standard deviation is larger than that of $\tilde{\beta}$. This is consistent with our observation after Theorem 1. The asymptotic expansion of $\hat{\beta} - \beta$ involves additional stochastic terms. Although their magnitudes are controlled asymptotically, their presence does influence the finite sample performance. $\check{\beta}$'s performance is generally best in terms of root mean squared error, followed by $\tilde{\beta}$ and $\tilde{\beta}^H$, $\check{\beta}^{(2)}$, and $\hat{\beta}$, with exceptions at $\theta = 0.2$ where $\tilde{\beta}$ performs best. $\check{\beta}$'s good performance is expected because with g_1, $\check{\beta}$ correctly assumes $E(X_t|\mathbf{Z_t})$ to be linear in $\mathbf{Z_t}$. When the heteroskedasticity assumption is in place, we notice the best estimator is $\tilde{\beta}^H$ in terms of bias, standard deviation, and root mean squared error. The reduction of standard deviation relative to $\tilde{\beta}$ is well over 10%. The observation confirms our theoretical result in Section 3 that $\tilde{\beta}^H$ properly takes into account the heteroskedasticity structure and reaches the semiparametric efficiency bound. The conclusion we draw for the performance of the rest of the estimators is similar to that in the homoskedasticity case, except that we notice $\check{\beta}^{(2)}$ performs much better, outperformed only by $\tilde{\beta}^H$ in terms of root mean squared error.

When g_2 and homoskedasticity are in the data-generating process, we note $E(X_t|\mathbf{Z_t})$ is not linear in $\mathbf{Z_t}$. $\check{\beta}^{(1)}$ carries the smallest standard deviation with a relatively large bias, while its root mean squared error is generally larger than $\tilde{\beta}$, $\tilde{\beta}^H$, and $\hat{\beta}$. Among the other estimators, $\tilde{\beta}$, $\tilde{\beta}^H$, and $\hat{\beta}$ are the best in terms of small bias and standard deviation, especially when the endogeneity problem is more severe. It is quite obvious that $\tilde{\beta}$ has the smallest root mean squared error. $\tilde{\beta}^H$ performs similarly well, followed by $\hat{\beta}$. $\check{\beta}^{(2)}$ and $\check{\beta}$ are outperformed by the others, due to the fact that $E(X_t|\mathbf{Z_t})$ is not linear in the data-generating process. This again confirms our results in Theorem 2. When the heteroskedasticity situation is considered, we notice that $\tilde{\beta}^H$ once again performs best among all estimators. The comments about the other estimators in the homoskedasticity case continue to be valid.

To summarize, under homoskedasticity, we conclude that when $E(X_t|\mathbf{Z_t})$ is linear in $\mathbf{Z_t}$, all estimators taking endogeneity into account perform better than $\check{\beta}^{(1)}$. $\check{\beta}$ is the best in this case with $\tilde{\beta}$, $\tilde{\beta}^H$, and $\check{\beta}^{(2)}$ being competitive alternatives. When $E(X_t|\mathbf{Z_t})$ is not linear in $\mathbf{Z_t}$, estimators $\tilde{\beta}$, $\tilde{\beta}^H$, and $\hat{\beta}$ perform best because they estimate $E(X_t|\mathbf{Z_t})$ nonparametrically, thus avoiding potential misspecification. We generally recommend the use of $\tilde{\beta}$ over $\hat{\beta}$ due to better finite sample performance. Under heteroskedasticity, however, the best estimator is $\tilde{\beta}^H$, which properly accounts for the information in the variance structure. $\tilde{\beta}$ continues to be a competitive alternative, though less efficient.

We further compare our estimators $\tilde{\beta}$ and $\tilde{\beta}^H$ with the efficient estimator $\tilde{\beta}^S$ proposed in Ai and Chen (2003) by a simple simulation analysis. The semiparametric efficient estimator $\tilde{\beta}^S$ is based on the method of minimum distance and sieves and can be applied to a more general semiparametric model. However, the efficient estimation calls for a numerical optimization. Since the estimators are constructed with different nonparametric methods, we follow their simulation closely for a meaningful comparison. The data generating process is the same as in their section 7, where the only change here is that we let the endogenous variable X_t shows up only in the parametric component, but not in the nonparametric control function. We investigate the performances with sample sizes $200, 400$ and perform 200 repetitions. We denote the degree of endogeneity by ρ (they use "R" for ρ. Here we use R to denote root mean squared error.) and consider $\rho = 0.1$ and 0.9, where a larger ρ indicates a more serious endogeneity.

The estimators' performances in terms of the bias, standard deviation and root mean squared error in the simulation are summarized in Table 5 in the Appendix. The results indicate $\tilde{\beta}$ and $\tilde{\beta}^H$ carry positive bias, while $\tilde{\beta}^S$ are negatively biased. All estimators' performance improve as the sample size increases. When endogeneity is more serious, it is more difficult for all to estimate β well, as judged by their larger R. Our estimators outperform $\tilde{\beta}^S$ in terms of standard deviation and root mean square error, and the advantage is more obvious in small samples. The limited simulation results suggest that with explicit algebraic structure and an easily implemented procedure without numerical optimization, our kernel-based estimators provide a viable alternative that complements the estimators available in the literature.

5 Empirical Illustration

We estimate the return to schooling with our proposed estimators using the data from Card (1995) in equation (1).[6] Using 3010 observations from the National Longitudinal Survey of Young Men in 1976, Card considers the regression model

$$\begin{array}{llll} logwage = \beta_0 & +\beta_1 educ & +\beta_2 exper & +\beta_3 exper^2/100 \\ OLS & 0.074 & 0.084 & -0.224 \\ 2SLS & 0.132 & 0.107 & -0.228 \\ & +\beta_4 black & +\beta_5 south & +\beta_6 smsa + \epsilon \\ & -0.190 & -0.125 & 0.161 \\ & -0.131 & -0.105 & 0.131 \end{array}$$

where *logwage* is the log of the 1976 hourly wage, *educ* is years of schooling, *exper* is the potential experience constructed as $age - educ - 6$. *black*, *south*, and *smsa* (Standard Metropolitan Statistical Area) are dummy variables (see Card for detailed data description and analysis). The OLS estimate for β_1 is 0.074 with a standard error of 0.004. However, due to the fact that *educ* is not randomly assigned or endogenous, it is difficult to argue it to be the return to schooling. Card uses the proximity to a four-year college (*nearc*4) as an instrumental variable for *educ*. Since being located close to a college might reduce the cost of investing in education, one might get more education. The 2SLS estimate for β_1 is 0.132 with a standard error of 0.049. Thus, the return to schooling might be underestimated in OLS[7], even though the OLS estimate carries a much smaller standard error. We notice the specification might be restrictive, i.e., experience enters the control function as a quadratic function, the dummy variables carry fixed coefficients and the reduced form is linear. The

[6] As in most of the work in the literature, we assume that the slope of the earning function does not vary across individuals, so that β in equation (1) can be interpreted as the return to schooling. An alternative framework would be to allow heterogeneous returns to schooling using a varying coefficient approach (see Su et al. (2011)). Deschenes (2007) estimates an interesting model of schooling and earnings with heterogeneous return to education. How our proposed estimators can be extended to such alternatives would be a topic for future research.

[7] 2SLS estimate could be biased in the same direction as OLS estimate in finite samples if the excluded instrumental variables explain a small share of the variation of the endogenous variables (Angrist and Krueger (1995).

larger value of β_1 in 2SLS might be due to the restrictive function form assumptions. We therefore consider the semiparametric estimation of β in the model

$$logwage = \beta educ + m(exper, black, south, smsa) + residuals,$$

using the two estimators proposed in the paper, $\tilde{\beta}$ and $\hat{\beta}$, where we allow $m(\cdot)$ and the reduced form to be nonparametric. Since the 2SLS results reported in Card are quite stable for treating experience as either exogenous or endogenous, we simply treat experience as an exogenous variable for illustration purposes. Using the proximity to a four-year college as the IV for *educ* as in Card (1995),[8] we choose the kernel function and bandwidths as in the Monte Carlo study. When $\tilde{\beta}$ is used, one can construct an estimate for $m(\cdot)$ as

$$\tilde{m}(Z_{1t}) = \hat{E}(Y|Z_{1t}) - \hat{E}(X|Z_{1t})\tilde{\beta}.$$

The estimated residual is $\tilde{\epsilon}_t = Y_t - X_t\tilde{\beta} - \tilde{m}(Z_{1t})$. Based on our Theorem 1, we can easily estimate the variance for $\tilde{\beta}$ as $(\hat{W}'\hat{W})^{-1}\frac{1}{n}\sum_{t=1}^{n}\tilde{\epsilon}_t^2$, if ϵ is conditionally homoskedastic. If not, the heteroskedasticity consistent variance estimate is constructed as $(\hat{W}'\hat{W})^{-1}\hat{W}\tilde{\Omega}\hat{W}'(\hat{W}'\hat{W})^{-1}$ following White (1980, 1982), where $\tilde{\Omega}$ is a diagonal matrix with $\tilde{\epsilon}_t^2$ as its diagonal term. A similar estimate could be obtained for the variance of $\hat{\beta}$.

We obtain $\tilde{\beta} = 0.048$ with a standard error of 0.019, and $\hat{\beta} = 0.091$ with a standard error of 0.019. The heteroskedasticity robust standard error is 0.016 for both estimates. It suggests that the return to schooling obtained in 2SLS might be too big due to its restrictive function form on the control variables and the reduced form. Our semiparametric efficient estimates are much more precise than the 2SLS estimates. To further illustrate the difference, we provide the estimated control function $m(exper, black, south, smsa)$ using $\tilde{\beta}$ and 2SLS in Figure 1 in the Appendix, for all combinations of the race and location status except for $(black, south, smsa) = (1, 0, 0)$ which has only five occurrences in the sample. Figure 1 shows that all $m(\cdot)$ estimated with 2SLS are quadratic

[8]Kling (2001) discussed some important and subtle issues in the choice of the instrumental variables for estimating the return to schooling.

in experience with different intercepts, with the difference between each of the two $m(\cdot)'s$ being fixed. On the other hand, our semiparametric estimates with $\tilde{\beta}$ give quite different results. For example, the quadratic control function assumption might be reasonable for a non-black person in the northern metropolitan area $((black, south, smsa) = (0, 0, 1))$, but it is unlikely to be the case for a black person located in the southern non-metropolitan area $((black, south, smsa) = (1, 1, 0))$. The gap between their expected wage is indeed largest in semiparametric estimates, but the gap changes across different experiences. Furthermore, the intercepts of $m(\cdot)$ estimated with $\tilde{\beta}$ are higher than those estimated with 2SLS, which follows because the $\tilde{\beta}$ estimate is smaller. To get an overall picture of the difference of two estimation procedures (semiparametric IV with $\tilde{\beta}$ and $2SLS$) in capturing *logwage*, we plot the predicted *logwage* against the realized *logwage* in Figure 2. A solid line is superimposed to indicate perfect prediction. Both estimates seem to overestimate *logwage* when it is small, but underestimate *logwage* when it is large. The semiparametic IV estimates' variability seems to be smaller.

Another potential source of misspecification is in the reduced form. For illustration purposes, we plot in Figure 3 the estimated nonparametric reduced form $E(educ|exper, black, south, smsa, nearc4)$ with $\tilde{\beta}$ and the estimated linear reduced form $\hat{educ} = \mathbf{Z_t}\hat{\alpha}$ used in 2SLS (note $\check{\beta}$ uses the same linear reduced form as 2SLS), where $\mathbf{Z_t} = (constant, exper, exper^2/100, black, south, smsa, nearc4)$, and $\hat{\alpha}$ is the linear regression parameter estimates. The 2SLS reduced forms of *educ* plotted against *exper* resemble straight lines with a fixed slope coefficient of -0.41, since the coefficient of the quadratic term $exper^2/100$, 0.073, is fairly small. As expected, the difference across different $(black, south, smsa, nearc4)$ groups are fixed. Residing close to a four-year college results in 0.337 more years in expected education, indicated by the larger intercept for reduced forms using 2SLS and $nearc4 = 1$ relative to $nearc4 = 0$. On the other hand, the nonparametric estimated reduced form delivers a quite different picture. Though one gets less *educ* as *exper* increases, the pattern is far from a linear one. The reduced form turns out to be approximately convex for education less than ten years, but roughly concave for education larger than ten years. Though one can say that a white person in a big city in the north gets more *educ* across almost all experience than other groups, the gaps of *educ* across different groups are far from a constant. For example, a black person from a small town in the south may not always get the least *educ*, and the ranking of the magnitude of *educ* across groups changes with experience.

Since 2SLS estimation is potentiallly misspecified in both the control function and the reduced form, we attempt to disentangle empirically the impact of misspecifications from the two sources, and assess which part contributes more to the difference in the return to schooling estimates obtained with $\tilde{\beta}$ and 2SLS. The estimator $\check{\beta}$ considered in equation (3) allows the control function $m(\cdot)$ to be a nonlinear function, but assumes the reduced form to be linear. We have demonstrated the potential efficiency gain of our estimator $\tilde{\beta}$ over $\check{\beta}$ in Theorem 2 and in the Monte Carlo study in situations where the reduced form is nonlinear. $\check{\beta}$ using $\mathbf{Z_t}\hat{\alpha}$ as the reduced form estimate turns out to be 0.135, with 0.052 as its standard error and the heteroskedasticity robust standard error estimate. Both the parameter and standard error estimates resemble those of 2SLS. It indicates that assuming the control function to be quadratic may not be far from the truth, but the parametric assumption on the reduced form is too restrictive. Thus, we conclude that further allowing nonparametric reduced form in $\tilde{\beta}$ not only allows us to construct efficient estimates relative to both $\check{\beta}$ and 2SLS, but also helps us to gain some empirical evidence that the difference between the return to schooling estimates obtained with $\tilde{\beta}$ and 2SLS largely comes from the misspecification of the reduced form.

Since the potential experience is closely related to education, *exper* might be endogenous for the same reason that education is endogenous. To check the robustness of the above findings, we consider the alternative model as $logwage = \beta educ + m(age, black, south, smsa) + residuals$, which replaces *exper* in the control function by *age*. We repeat the semiparametric IV estimation procedures with proximity to a four-year college as IV for *educ*. We obtain $\tilde{\beta} = 0.024$ with a standard error of 0.019 and $\hat{\beta} = 0.039$ with a standard error of 0.019. The heteroskedasticity robust standard error is 0.017 for both estimates. Again, the semiparametric efficient estimates are much more precise than the 2SLS estimate. Although the estimates change, we still conclude that the 2SLS estimate for the return to schooling might be too large, due to its restrictive functional form assumptions on the control variables and on the reduced form.

6 Conclusion

We provide an explanation for the puzzle observed in Card (2001) that 2SLS estimates for the return to schooling are typically larger than OLS estimates, though the former are fairly imprecisely estimated. The difference could be due to the restrictive linear functional form specification in 2SLS on the control variables and the reduced form. For the parameters of endogenous regressors, we propose three kernel-based semiparametric IV estimators that relax the tight functional form assumptions on the control variables and the reduced form. They have explicit algebraic structures and are easily implemented without numerical optimizations. We show that they are consistent and asymptotically normal. The first two estimators are efficient relative to previously considered estimators under homoskedasticity. The third estimator incorporates heteroskedasticity information and is efficient under heteroskedasticity. They reach the semiparametric efficiency bounds in Chamberlain (1992), and are asymptotically equivalent to semiparametric IV estimators that optimally select the instrument under conditional moment restrictions. A Monte Carlo study shows that they perform well in finite samples. We estimate the return to schooling with the proposed estimators using data in Card (1995). We find that the estimate for the return to schooling is much smaller and more precise than the 2SLS estimate and the difference largely comes from the misspecification in the linear reduced form.

Appendix

Table 1 Bias($\times$0.1)(B), Standard deviation(S) and Root Mean Squared Error(R) for estimators with g_1 and σ_1^2

	$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
$n = 100$	B	S	R	B	S	R	B	S	R
$\breve{\beta}^{(1)}$	.963	.073	.121	2.469	.066	.256	3.947	.052	.398
$\breve{\beta}^{(2)}$	.050	.119	.119	.062	.124	.124	-.104	.127	.128
$\breve{\beta}$	-.022	.107	.107	-.011	.110	.110	-.109	.113	.113
$\hat{\beta}$	.879	.107	.139	1.360	.108	.174	1.789	.104	.207
$\tilde{\beta}$	.262	.095	.099	.663	.094	.115	.991	.088	.133
$\tilde{\beta}^H$	.266	.101	.105	.685	.100	.121	1.030	.093	.139
$n = 200$	B	S	R	B	S	R	B	S	R
$\breve{\beta}^{(1)}$	.988	.050	.111	2.461	.047	.251	3.959	.035	.397
$\breve{\beta}^{(2)}$	.051	.079	.079	-.007	.082	.082	-.021	.079	.079
$\breve{\beta}$	-.031	.073	.073	-.060	.074	.074	-.076	.071	.072
$\hat{\beta}$	.768	.073	.106	1.117	.075	.134	1.448	.069	.161
$\tilde{\beta}$	.159	.068	.070	.404	.067	.079	.627	.062	.088
$\tilde{\beta}^H$	.160	.070	.072	.413	.069	.081	.644	.064	.091

Table 2 Bias(B), Standard deviation(S) and Root Mean Squared Error(R) for estimators with g_2 and σ_1^2

	$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
$n = 100$	B	S	R	B	S	R	B	S	R
$\breve{\beta}^{(1)}$	.152	.091	.177	.384	.079	.392	.630	.058	.632
$\breve{\beta}^{(2)}$	.525	10.177	10.185	.294	9.011	9.011	.285	24.627	24.617
$\breve{\beta}$	.182	3.589	3.592	.178	12.492	12.487	.532	1.906	1.978
$\hat{\beta}$	.215	.209	.300	.353	.205	.409	.504	.215	.548
$\tilde{\beta}$	.093	.152	.178	.204	.141	.248	.333	.118	.353
$\tilde{\beta}^H$	.099	.160	.188	.214	.150	.261	.342	.126	.364
$n = 200$	B	S	R	B	S	R	B	S	R
$\breve{\beta}^{(1)}$	.155	.066	.168	.390	.051	.393	.627	.039	.628
$\breve{\beta}^{(2)}$	.629	20.838	20.838	3.628	57.572	57.657	4.768	157.262	157.255
$\breve{\beta}$	.183	3.215	3.218	.393	3.078	3.101	.599	2.416	2.488
$\hat{\beta}$	.221	.151	.268	.325	.146	.356	.436	.150	.461
$\tilde{\beta}$	.073	.118	.138	.156	.108	.189	.237	.092	.255
$\tilde{\beta}^H$	.076	.121	.143	.160	.111	.194	.241	.093	.259

Table 3 Bias($\times 0.1$)(B), Standard deviation(S) and Root Mean Squared Error(R) for estimators with g_1 and σ_2^2

	$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
$n = 100$	B	S	R	B	S	R	B	S	R
$\check{\beta}^{(1)}$	.345	.058	.067	1.028	.054	.116	1.674	.052	.175
$\check{\beta}^{(2)}$	.033	.054	.054	.049	.054	.054	-.005	.057	.057
$\check{\beta}$	-.124	.060	.062	-.096	.059	.060	-.120	.062	.063
$\hat{\beta}$	.376	.062	.073	.594	.062	.086	.741	.064	.098
$\tilde{\beta}$	.038	.058	.058	.221	.056	.060	.346	.056	.066
$\tilde{\beta}^H$	.012	.038	.038	.076	.037	.038	.106	.038	.040
$n = 200$	B	S	R	B	S	R	B	S	R
$\check{\beta}^{(1)}$	.385	.039	.055	1.051	.037	.111	1.719	.035	.175
$\check{\beta}^{(2)}$	.045	.035	.035	.022	.037	.037	.030	.036	.036
$\check{\beta}$	-.071	.039	.039	-.095	.042	.043	-.086	.041	.042
$\hat{\beta}$	.406	.041	.058	.509	.044	.067	.655	.043	.078
$\tilde{\beta}$	.033	.039	.039	.105	.041	.042	.209	.039	.044
$\tilde{\beta}^H$	.003	.024	.024	.017	.024	.024	.033	.024	.024

Table 4 Bias(B), Standard deviation(S) and Root Mean Squared Error(R) for estimators with g_2 and σ_2^2

	$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
$n = 100$	B	S	R	B	S	R	B	S	R
$\check{\beta}^{(1)}$	.097	.093	.134	.242	.084	.256	.397	.069	.403
$\check{\beta}^{(2)}$	.208	10.796	10.792	.353	8.366	8.370	.844	13.184	13.204
$\check{\beta}$	.076	2.955	2.954	.227	2.620	2.629	.197	2.703	2.709
$\hat{\beta}$	.115	.143	.183	.177	.141	.226	.242	.141	.280
$\tilde{\beta}$	.063	.107	.125	.112	.100	.151	.170	.092	.194
$\tilde{\beta}^H$	.016	.075	.076	.037	.079	.087	.063	.082	.104
$n = 200$	B	S	R	B	S	R	B	S	R
$\check{\beta}^{(1)}$	.093	.065	.113	.251	.055	.257	.407	.043	.409
$\check{\beta}^{(2)}$	-.036	24.853	24.840	.199	16.715	16.708	.302	6.619	6.623
$\check{\beta}$	.131	2.197	2.200	.155	2.211	2.215	.341	1.546	1.582
$\hat{\beta}$	.118	.095	.151	.172	.094	.196	.219	.093	.238
$\tilde{\beta}$	.042	.071	.083	.085	.071	.111	.123	.066	.140
$\tilde{\beta}^H$	.075	.044	.045	.015	.045	.047	.030	.044	.053

Table 5 Bias(B), Standard deviation(S) and Root Mean Squared Error(R) for estimators $\tilde{\beta}$, $\tilde{\beta}^H$ and $\tilde{\beta}^S$

	$\rho = 0.1$						$\rho = 0.9$					
	$n = 200$			$n = 400$			$n = 200$			$n = 400$		
	B	S	R	B	S	R	B	S	R	B	S	R
$\tilde{\beta}$	.015	.113	.114	.026	.110	.113	.239	.097	.257	.197	.078	.212
$\tilde{\beta}^H$	.011	.106	.106	.019	.099	.101	.212	.098	.233	.172	.079	.189
$\tilde{\beta}^S$	-.057	.563	.565	.017	.222	.222	-.126	1.181	1.185	-.025	.247	.247

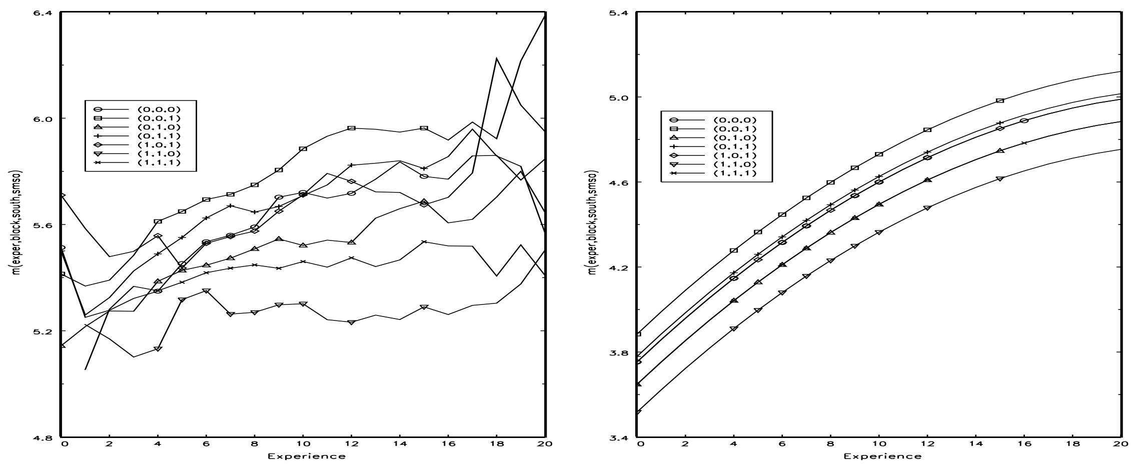

Figure 1: Plot of estimated control function $m(exper, black, south, smsa)$ with $\tilde{\beta}$ (left) and $2SLS$ (right). $(black, south, smsa)$ indicated in graph.

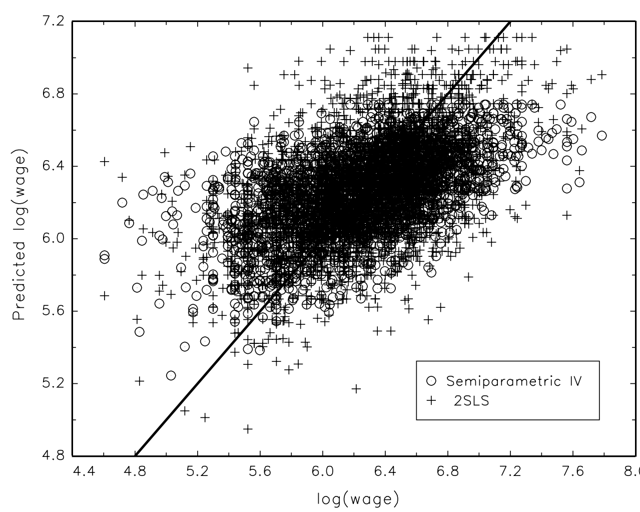

Figure 2: Plot of estimated $log(wage)$ using $\tilde{\beta}$ and $2SLS$ against realized $log(wage)$.

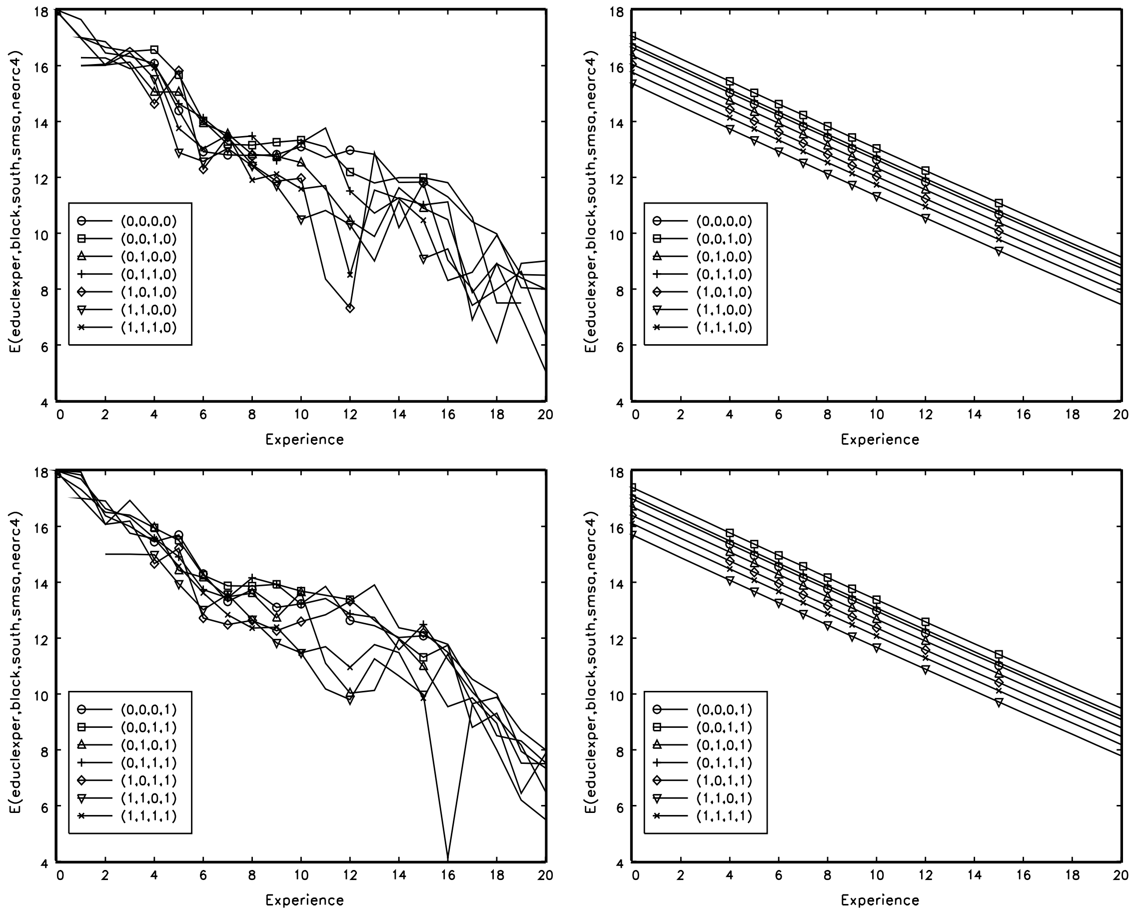

Figure 3: Plot of estimated reduced form $E(educ|exper, black, south, smsa, nearc4)$ with $\tilde{\beta}$ (left) and $2SLS$ (right). $(black, south, smsa, nearc4)$ indicated in graph. $neac4 = 0$ for the top panels.

Appendix of proof

Lemma 1 *Define*

$$S_{n,j}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i^c - z^c)\left(\frac{Z_i^c - z^c}{h}\right)^j I(Z_i^d = z^d)g(U_i)w(Z_i^c - z^c; z), |j| = 0, 1, 2, \cdots, J,$$

where Z_i, U_i *are iid,* $Z_i^c \in R^{l_c}, Z_i^d \in R^{l_d}, K_h(z^c) = \frac{1}{h^{l_c}}K(\frac{z^c}{h})$*, and* $K(.)$ *is a kernel function defined on* R^{l_c}*. If we have*

L_1. $K(.)$ *is bounded with compact support and for Euclidean norm* $||.||$,

$$|u^j K(u) - v^j K(v)| \leq c_K ||u - v||, \text{ for } 0 \leq |j| \leq J.$$

L_2. $g(u)$ *is a measurable function of* u_i *and* $E|g(u)|^s < \infty$ *for* $s > 2$.
L_3. $\sup_{z\in G}\int |g(u)|^s f_{z,u}(z,u)du < \infty$, $f_{z|u}(z) < \infty$, *and* $f_{z,u}(z,u)$ *is continuous around* z^c.
L_4. $|w(Z_i^c - z^c; z)| < \infty$, $\forall z^d \in G^d$, *a compact subset of* R^{l_d}, $|w(Z_i^c - z^c; z^c, z^d) - w(Z_i^c - z_k^c; z_k^c, z^d)| \leq c||z^c - z_k^c||$.
L_5. $nh^{l_c} \to \infty$.

Then for $z = (z^c, z^d) \in G = G^c \times G^d$, $z^c \in G^c$, a compact subset of R^{l_c},

$$\sup_{z\in G} |S_{n,j}(z) - E(S_{n,j}(z))| = O_p\left(\left(\frac{nh^{l_c}}{ln(n)}\right)^{-\frac{1}{2}}\right).$$

Proof. Let's define

$$S^B_{n,j}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(Z_i^c - z^c)\left(\frac{Z_i^c - z^c}{h}\right)^j I(Z_i^d = z^d) g(U_i) w(Z_i^c - z^c; z) I(|g(U_i)| \le B_n),$$

where $B_1 \le B_2 \le \cdots$ such that $\sum_{i=1}^{\infty} B_i^{-s} < \infty$ for some $s > 0$. Since $G^c \times G^d$ is compact, we could cover G by a finite number l_n of l_c dimensional cubes I_k with center z_k, $k = 1, 2, \cdots, l_n$ and length r_n. We could choose l_n sufficiently large such that r_n is sufficiently small and each cube I_k corresponds to one fixed possible value of z^d, i.e., $z^d = z^d_k$ if $z \in I_k$. Since G is compact, $l_n r_n^{l_c} = c$, c a constant. Suppose we let $l_n = \left(\frac{n}{ln(n)h^{l_c+2}}\right)^{\frac{l_c}{2}}$, then $r_n = c/l_n^{\frac{1}{l_c}}$. Since

$$\begin{aligned}
& \sup_{z\in G} |S^B_{n,j}(z) - E(S^B_{n,j}(z))| \\
= \; & max_{1\le k\le l_n} \sup_{z\in I_k\cap G} |S^B_{n,j}(z) - E(S^B_{n,j}(z)) \\
= \; & max_{1\le k\le l_n} \sup_{z^c\in I_k\cap G} |S^B_{n,j}(z^c, z^d_k) - S^B_{n,j}(z_k) \\
& + S^B_{n,j}(z_k) - ES^B_{n,j}(z_k) + ES^B_{n,j}(z_k) - ES^B_{n,j}(z^c, z^d_k)| \\
\le \; & max_{1\le k\le l_n} \sup_{z^c\in I_k\cap G} |S^B_{n,j}(z^c, z^d_k) - S^B_{n,j}(z_k)| \\
& + max_{1\le k\le l_n} |S^B_{n,j}(z_k) - ES^B_{n,j}(z_k)| \\
& + max_{1\le k\le l_n} \sup_{z^c\in I_k\cap G} |ES^B_{n,j}(z_k) - ES^B_{n,j}(z^c, z^d_k)| \\
= \; & I_1 + I_2 + I_3
\end{aligned}$$

The lemma is proved if we can show

(1) $I_0 = \sup_{z\in G} |S_{n,j}(z) - E(S_{n,j}(z)) - [S^B_{n,j}(z) - E(S^B_{n,j}(z))]| = O_{a.s.}(B_n^{1-s})$ for $B_n^{1-s} = O(\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}})$, $\sum_{i=1}^{\infty} B_i^{-s} < \infty$.

(2) $I_1 = O_{a.s.}(\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}})$. (3) $I_2 = O_p(\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}})$. (4) $I_3 = O_{a.s.}(\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}})$.

(1) $I_0 \le \sup_{z\in G} |S_{n,j}(z) - S^B_{n,j}(z)| + \sup_{z\in G} |E(S_{n,j}(z) - S^B_{n,j}(z))| = I_{01} + I_{02}$. We note

$$I_{01} = \sum_{z\in G} |\frac{1}{n}\sum_{i=1}^{n} K_h(Z_i^c - z^c)\left(\frac{Z_i^c - z^c}{h}\right)^j I(Z_i^d = z^d) g(U_i) w(Z_i) I(|g(U_i)| > B_n).$$

BY Chebychev's inequality, $\sum_{i=1}^{\infty} P(|g(U_i)| > B_i) \le \sum_{i=1}^{\infty} \frac{E|(g(U_i)|^s}{B_i^s} < c\sum_{i=1}^{\infty} B_i^{-s} \le \infty$, by construction of B_i and L_2. By Borel-Cantellis Lemma, $P(|g(U_i)| > B_i \quad i.o.) = 0$. To see this, $P(|g(U_i)| > B_i \quad i.o.) = \lim_{i\to\infty} P(\cup_{m=i}^{\infty}\{\omega : |g(U_m)| > B_m\}) \le \lim_{i\to\infty}\sum_{m=i}^{\infty} P(\{\omega : |g(U_m)| > B_m\}) = 0$ since $\sum_{i=1}^{\infty} P(\{\omega : |g(U_i)| > B_i\}) < \infty$. So $\forall\epsilon > 0$, there exists $i' > 0$ such that $\forall i > i'$,

$$P(\cup_{m=i}^{\infty}\{\omega : |g(U_m)| > B_m\}) < \epsilon, \quad \text{or } P(\cap_{m=i}^{\infty}\{\omega : |g(U_m)| \le B_m\}) > 1 - \epsilon.$$

So $\forall m > i'$, $P(|g(U_m)| \le B_m) > 1 - \epsilon$ or $|g(U_m)| \le B_m$ for sufficiently large m. Since B_i is an increasing sequence, w.p.1, $|g(U_m)| \le B_n$ for $m \ge i'$ and $n \ge m$.

When $i = \{1, 2, \cdots, i'\}$, $P(|g(U_i)| \le B_n) > 1 - \epsilon$. To see this, $\forall\epsilon > 0$, and sufficiently large n, $P(|g(U_i)| > B_n) < \frac{E|g(U_i)|^s}{B_n^s} < \frac{c}{B_n^s} < \epsilon$, since $E|g(U_i)|^s < \infty$ and B_i is an increasing sequence. So in all, $\forall\epsilon > 0$, and for n sufficiently large, we have $I(|g(U_i)| > B_n) = 0$ w.p.1.. So $I_{01} = 0$ a.s..

$I_{02} = \sup_{z\in G} |EK_h(Z_i^c - z^c)\left(\frac{Z_i^c - z^c}{h}\right)^j I(Z_i^d = z^d) g(U_i) w(Z_i^c - z^c; z) I(|g(U_i)| > B_n)$. Let $\frac{Z_i^c - z^c}{h} = (\frac{Z_{i1}^c - z_1^c}{h}, \cdots, \frac{Z_{il_c}^c - z_{l_c}^c}{h}) = \Psi_i = (\Psi_{i1}, \cdots, \Psi_{il_c})$, so $z_i^c = z_1^c + h\Psi_{i1}, \cdots, z_{l_c}^c + h\Psi_{il_c} = z^c + h\Psi_i$,

$|\frac{\partial z_i^c}{\partial \Psi_i}| = h^{l_c}$. By change of variable,

$$
\begin{aligned}
I_{02} = \;& \sup_{z\in G} |\textstyle\sum_{Z_i^d=z^d} \int K(\Psi_i)\Psi^j \int w(h\Psi_i; z) g(U_i) I(|g(U_i)| > B_n) \\
& \times f_{z,u}(z^c + h\Psi_i, z^d, U_i) dU_i d\Psi_i| \\
\leq \;& c \textstyle\int |K(\Psi_i)\Psi_i^j| d\Psi_i \sup_{z\in G} \int |g(U_i)| f_{z,U_i}(z, U_i) I(|g(U_i)| > B_n) dU_i \\
\leq \;& c \sup_{z\in G} [\textstyle\int |g(U_i)|^s f_{z,U_i}(z, U_i) dU_i]^{\frac{1}{s}} [\int I(|g(U_i)| > B_n) f_{z,U_i}(z, U_i) dU_i]^{1-\frac{1}{s}} \\
\leq \;& c[E_{U_i}(I(|g(U_i)| > B_n) f_{z|U_i}(z))]^{1-\frac{1}{s}} \\
\leq \;& c[E_{U_i}(I(|g(U_i)| > B_n)]^{1-\frac{1}{s}} = c[P(|g(U_i)| > B_n)]^{1-\frac{1}{s}} \\
\leq \;& c[\tfrac{E|g(U_i)|^s}{B_n^s}]^{1-\frac{1}{s}} \leq c B_n^{1-\frac{1}{s}}.
\end{aligned}
$$

where to obtain the first inequality we use L_4, the second we use L_1 and Hölder's inequality, the third and fourth we use L_3. The last tine above we use Chebychev's inequality again and L_2.

$$
\begin{aligned}
& (2) \quad |S_{n,j}^B(z^c, z_k^d) - S_{n,j}^B(z_k)| \\
= \;& |\tfrac{1}{nh^{l_c}} \textstyle\sum_i [K_h(Z_i^c - z^c) \left(\frac{Z_i^c - z^c}{h}\right)^j w(Z_i^c - z^c; z^c, z_k^d) \\
& - K_h(Z_i^c - Z_k^c) \left(\frac{Z_i^c - Z_k^c}{h}\right)^j w(Z_i^c - z_k^c; z_k)] I(Z_i^d = Z_k^d) g(U_i) I(|g(U_i)| \leq B_n)| \\
\leq \;& \tfrac{1}{nh^{l_c}} \textstyle\sum_i [|[K(Z_i^c - z^c) \left(\frac{Z_i^c - z^c}{h}\right)^j - K(Z_i^c - z_k^c) \left(\frac{Z_i^c - z_k^c}{h}\right)^j] w(Z_i^c - z^c; z^c, z_k^d)| \\
& + |K(Z_i^c - Z_k^c) \left(\frac{Z_i^c - Z_k^c}{h}\right)^j [w(Z_i^c - z^c; z^c, z_k^d) - w(Z_i^c - z_k^c; z_k)]|] \\
& \times I(Z_i^d = Z_k^d) g(U_i) I(|g(U_i)| \leq B_n)| \\
\leq \;& \tfrac{1}{nh^{l_c}} \textstyle\sum_i [c \frac{||Z_k^c - z^c||}{h} + c||Z_k^c - z^c||]|g(U_i)| \text{ by } L_1 \text{ and } L_4,
\end{aligned}
$$

since $z \in I_k$ for some k, $||Z_k^c - z^c|| \leq c r_n$ and with L_4,
$I_1 \leq c\frac{r_n}{h^{l_c+1}} \frac{1}{n} \sum_i |g(U_i)|$, by L_2 and Kolmogorov's Theorem,

$$\frac{1}{n} \sum_i |g(U_i)| \stackrel{a.s.}{\to} E|g(U_i)| < \infty.$$

So $I_1 \leq c\frac{r_n}{h^{l_c+1}} = \frac{c}{h^{l_c+1}} (\frac{n}{ln(n)h^{l_c+2}})^{-\frac{1}{2}} = c(\frac{ln(n)}{nh^{l_c}})^{\frac{1}{2}}$ *a.s.*.

We could show (4) $I_3 = O_{a.s.}\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}}$ similarly.

(3) It is sufficient to show $\exists$ a constant $\Delta > 0$ and $N > 0$ such that $\forall \epsilon > 0$ and $n > N$, $P(\left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}} I_2 \geq \Delta) < \epsilon$.

Let $\epsilon_n = \left(\frac{ln(n)}{nh^{l_c}}\right)^{\frac{1}{2}} \Delta$, then $P(I_2 \geq \epsilon_n) \leq \sum_{k=1}^{l_n} P(|S_{n,j}^B(z_k) - ES_{n,j}^B(z_k)| \geq \epsilon_n)$. We note $|S_{n,j}^B(z_k) - ES_{n,j}^B(z_k)|$
$= |\frac{1}{n} \sum_{i=1}^n [\frac{1}{h^{l_c}} K(Z_i^c - z_k^c) \left(\frac{Z_i^c - z_k^c}{h}\right)^j I(Z_i^d = z_k^d) g(U_i) w(Z_i^c - z_k^c; z_k) I(|g(U_i)| \leq B_n) - \frac{1}{h^{l_c}} EK(Z_i^c - z_k^c) \left(\frac{Z_i^c - z_k^c}{h}\right)^j I(Z_i^d = z_k^d) g(U_i) w(Z_i^c - z_k^c; z_k) I(|g(U_i)| \leq B_n)]| = |\frac{1}{n} \sum_{i=1}^n W_{in}|$.

Since $EW_{in} = 0$, $|W_{in}| \leq 2c\frac{B_n}{h^{l_c}}$ by L_1 and L_4, and $\{W_{in}\}_{i=1}^n$ is an independent sequence, by Bernstein's inequality,

$P(|S_{n,j}^B(z_k) - ES_{n,j}^B(z_k)| \geq \epsilon_n) < 2exp\left(\frac{-nh^{l_c}\epsilon_n^2}{2h^{l_c}\bar{\sigma}^2 + \frac{2}{3}B_n\epsilon_n}\right)$,

where $\bar{\sigma}^2 = \frac{1}{n} \sum_i V(W_{in}) = EW_{in}^2 = I_{21} - I_{22}^2$

$$= \frac{1}{h^{2l_c}} EK^2(Z_i^c - z^c)(\left(\frac{Z_i^c - z_k^c}{h}\right)^j)^2 I(Z_i^d = z_k^d) g^2(U_i) w(Z_i^c - z_k^c; z_k)^2 I(|g(U_i)| \leq B_n)]|$$

$- \left[\frac{1}{h^{l_c}} EK(Z_i^c - z_k^c) \left(\frac{Z_i^c - z_k^c}{h} \right)^j I(Z_i^d = z_k^d) g(U_i) w(Z_i^c - z_k^c; z_k) I(|g(U_i)| \le B_n) \right] \right]^2.$

$I_{22} = \sum_{Z_i^d = z_k^d} \int K(\Psi)\Psi_i^j g(U_i) w(h\Psi_i; z_k) I(|g(U_i)| \le B_n) f_{z,u}(z_k^c + h\Psi_i, Z_i^d, U_i) d\psi_i dU_i$

$\le c \int |K(\Psi)\Psi_i^j||g(U_i)| f_{z,u}(Z_k^c + h\Psi_i, Z_k^d, U_i) d\psi_i dU_i$

$\to c \int |K(\Psi)\Psi_i^j| d\Psi_i \int |g(U_i)| f_{z,u}(z_k, U_i) dU_i < \infty$, with L_1, L_3 and L_4.

Similarly $h^{L_c} I_{21} = O(1)$. So $2h^{l_c}\bar{\sigma}^2 < \infty$. If $B_n\epsilon_n < \infty$, then $C_n = 2h^{l_c}\bar{\sigma}^2 + \frac{2}{3}B_n\epsilon_n < \infty$, then $P(I_2 \ge \epsilon_n) \le l_n 2exp\left(\frac{-nh^{l_c}\epsilon_n^2}{2h^{l_c}\bar{\sigma}^2 + \frac{2}{3}B_n\epsilon_n} \right)$

$$= \left(\frac{n}{ln(n)h^{l_c+2}} \right)^{\frac{l_c}{2}} 2exp\left(\frac{-nh^{l_c}\left(\left(\frac{ln(n)}{nh^{l_c}} \right)^{\frac{1}{2}} \Delta \right)^2}{C_n} \right) = \frac{2n^{\frac{l_c}{2} - \frac{\Delta^2}{C_n}}}{(ln(n))^{\frac{l_c}{2}} h^{l_c + \frac{l_c^2}{2}}} \to 0.$$

Above is true since $C_n < \infty$, if we let $\Delta^2 \ge C_n(1 + l_c)$, then $\frac{2n^{\frac{l_c}{2} - \frac{\Delta^2}{C_n}}}{(ln(n))^{\frac{l_c}{2}} h^{l_c + \frac{l_c^2}{2}}} \le \frac{2}{(ln(n))^{\frac{l_c}{2}} (nh^{l_c})^{1 + \frac{l_c}{2}}} \to 0$ by L_5.

If we let $B_n = n^{\frac{1}{s}+\delta}$ for $s > 2$ and $\delta > 0$, then $B_n\epsilon_n < \infty$ for sufficiently large s. To see this, $B_n\epsilon_n = n^{\frac{1}{s}+\delta}\Delta(\frac{ln(n)}{nh^{l_c}})^{\frac{1}{2}}$. By L_5, we could let $(nh^{l_c})^{-\frac{1}{2}} = n^{-\frac{1}{2}+\delta_1}$ for $\frac{1}{2} > \delta_1 > 0$, then $B_n\epsilon_n = n^{\frac{1}{s} - \frac{1}{2} + \delta + \delta_1}\Delta(ln(n))^{\frac{1}{2}}$. If we let $s > [\frac{1}{2} - \delta - \delta_1]^{-1}$, then $B_n\epsilon_n \to 0$.

It is easy to see that for $B_n = n^{\frac{1}{s}+\delta}$, we easily have $\sum_{i=1}^{\infty} B_i^{-s} < \infty$. Furthermore $B_n^{1-s} < n^{\frac{1}{2}-\delta}$, so $B_n^{1-s} = O\left(\frac{ln(n)}{nh^{l_c}} \right)^{\frac{1}{2}}$.

Theorem 1: *Proof.* Note $\hat{E}(Y|Z_t) - \hat{E}(Y|Z_{1t}) = \hat{W}_t\beta + \hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + \hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})$, so we could write

$$\tilde{\beta} - \beta = [(\frac{1}{n}\hat{W}'\hat{W})^{-1} - (EW_t'W_t)^{-1} + (EW_t'W_t)^{-1}]$$
$$\times \underbrace{\frac{1}{n}\hat{W}'(\hat{E}(m(z_1)|\vec{Z}) - \hat{E}(m(z_1)|\vec{Z}_1) + \hat{E}(\epsilon|\vec{Z}) - \hat{E}(\epsilon|\vec{Z}_1))}_{C}.$$

Let's denote $\hat{E}(X_k|Z_t) = \hat{g}_k(Z_t)$ and $\hat{E}(X_k|Z_{1t}) = \hat{g}_{1,k}(Z_{1t})$, then $\hat{W}_{t,k} = \hat{g}_k(Z_t) - g_k(Z_t) + g_{1,k}(Z_{1t}) - \hat{g}_{1,k}(Z_{1t}) + g_k(Z_t) - g_{1,k}(Z_{1t})$, then $(i,j)th$ element of $\frac{1}{n}\hat{W}'\hat{W}$ is

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{t=1}^n \hat{W}_{t,i}\hat{W}_{t,j} \\
= \ & \tfrac{1}{n}\textstyle\sum_t [\hat{g}_i(Z_t) - g_i(Z_t)][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& + \tfrac{1}{n}\textstyle\sum_t [\hat{g}_i(Z_t) - g_i(Z_t)][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [\hat{g}_i(Z_t) - g_i(Z_t)][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_i(Z_t) - g_{1,i}(Z_{1t})][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_i(Z_t) - g_{1,i}(Z_{1t})][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_i(Z_t) - g_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
= \ & A_1 + A_2 + \cdots + A_9
\end{aligned}$$

Similarly, for $k = 1, 2, \cdots, K$, the kth element of C is

$$\begin{aligned}
C_k = \ & \tfrac{1}{n}\textstyle\sum_{t=1}^n \hat{W}_{t,k}(\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + \hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})) \\
= \ & \tfrac{1}{n}\textstyle\sum_t [\hat{g}_k(Z_t) - g_k(Z_t)][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_{1,k}(Z_{1t}) - \hat{g}_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_k(Z_t) - g_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [\hat{g}_k(Z_t) - g_k(Z_t)][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_{1,k}(Z_{1t}) - \hat{g}_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
& + \tfrac{1}{n}\textstyle\sum_t [g_k(Z_t) - g_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
= \ & C_{1k} + C_{2k} + \cdots + C_{6k}
\end{aligned}$$

We show below (1) $A_i = o_p(1)$, $i = 1, \cdots, 8$,

$$A_9 - E[g_i(Z_t) - g_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] = o_p(1),$$

so together we have $\frac{1}{n}\hat{W}'\hat{W} - EW_t'W_t = o_p(1)$. By $A1(3)$ and Slutsky' Theorem, $(\frac{1}{n}\hat{W}'\hat{W})^{-1} - (EW_t'W_t)^{-1} = o_p(1)$.
(2) $C_{1k} = [O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p((\frac{nh_1^{l_{1c}-2}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_2^{l_{1c}+l_{2c}-2}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1}) + O(h_2^{s_2+1})]$.
$C_{2k} = [O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})][O_p((\frac{nh_1^{l_{1c}-2}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_2^{l_{1c}+l_{2c}-2}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1}) + O(h_2^{s_2+1})]$.
$C_{3k} = O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1}) + O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-1})$
$+O_p(h_1^{s+1}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O((n^2h_1^{l_{1c}-2})^{-\frac{1}{2}}) + O(h_1^{s+1}) + O((n^2h_2^{l_{1c}+l_{2c}-2})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$.
$C_{4k} = [O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})]$.
$C_{5k} = [O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})][[O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})]$.
For $C_6 = [C_{61}, C_{62}, \cdots, C_{6K}]'$, $\sqrt{n}C_6 \xrightarrow{d} N(0, \Phi_0)$, where Φ_0 is defined in Theorem 1.
Since $\sqrt{n}C_{1k} = [O_p((\frac{(nh_2^{2(l_{1c}+l_{2c})})^{\frac{1}{2}}}{ln(n)})^{-\frac{1}{2}}) + O(n^{\frac{1}{4}}h_2^{s_1+1})]$
$\times[O_p(h_1(\frac{(nh_1^{2l_{1c}})^{\frac{1}{2}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_2(\frac{(nh_2^{2(l_{1c}+l_{2c})})^{\frac{1}{2}}}{ln(n)})^{-\frac{1}{2}}) + O(n^{\frac{1}{4}}h_1^{s+1}) + O(n^{\frac{1}{4}}h_2^{s_1+1})] = o_p(1)$ with $A5$.
Similar arguments could be used with A5 to show $\sqrt{n}C_{ik} = o_p(1)$ for $i = 2, 3, 4, 5$. Note the relatively strong assumption A5(3) are used specifically in C_{3k} to make the bias disappear asymptotically.
Combining results in (1) and (2) and using A1(3), we conclude

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{d} N(0, (EW_t'W_t)^{-1}\Phi_0(EW_t'W_t)^{-1}).$$

(1) (a) We first show $\sup_{z_{10}\in G_1} |\hat{f}_1(z_{10}) - f_1(z_{10})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^s)$.
We apply Lemma 1 with $S_{n,0}(z_{10}) = \frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})I(Z_{1i}^d = z_{10}^d)$, so

$$\sup_{z_{10}\in G_1} |\hat{f}_1(z_{10}) - E\hat{f}_1(z_{10})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}).$$

Condition L_1 is satisfied with A3, L_2 is satisfied since $g(u) = 1$, L_3 is true with A2(1) and (2), L_4 is satisfied since $w(z) = 1$. Since the data are iid in A1(1),
$E\hat{f}(z_{10}) = \int \frac{1}{h_1^{l_{1c}}}K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})f_1(Z_{1i}^c, z_{10}^d)dZ_{1i}^c$, with $\Psi_i = \frac{Z_{1i}^c - z_{10}^c}{h_1}$,
$= \int K_1(\Psi_i)f_1(z_{10}^c + h_1\Psi_i, z_{10}^d)d\Psi_i$, with A2(1)
$= \int K_1(\Psi_i)[f_1(z_{10}^c, z_{10}^d) + \sum_{|j|=1}^{s-1}\frac{h_1^{|j|}}{j!}\frac{\partial^j f_1(z_{10}^c, z_{10}^d)}{\partial(z_{10}^c)^j}\Psi_i^j + \sum_{|j|=s}\frac{h_1^s}{j!}\frac{\partial^j f_1(z_{10*}^c, z_{10}^d)}{\partial(z_{10}^c)^j}\Psi_i^j]d\Psi_i$
$= f_1(z_{10}) + O(h_1^s)$ uniformly $\forall z_{10} \in G_1$ by A3, A2(1) and Dominated Convergence Theorem, where z_{10*}^c is between z_{10}^c and z_{1i*}^c. So $\sup_{z_{10}\in G_1} |E\hat{f}_1(z_{10}) - f_1(z_{10})| = O(h_1^s)$.

(b) Similarly, we obtain $\sup_{z_0\in G} |\hat{f}(z_0) - f_z(z_0)| = O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1})$ with A2(4), (5) and A3.

(c) We show $\sup_{z_{10}\in G_1} |\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})I(Z_{1i}^d = z_{10}^d)[X_{i,k} - g_{1,k}(z_{10})]| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})$.
Since $X_{i,k} = g_{1,k}(Z_{1i}) + e_{1,ki}$, we have
$\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})I(Z_{1i}^d = z_{10}^d)e_{1,ki}$
$+\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})I(Z_{1i}^d = z_{10}^d)[g_{1,k}(Z_{1i}) - g_{1,k}(z_{10})] = I_1 + I_2$.

We apply Lemma 1 again with $S_{n,0}(z_{10}) = I_1$, $g(U_i) = e_{1,ki}$, and $w(z) = 1$. L_2 is implied by A4(1) and L_3 is implied by A2(3). Since $E(e_{1,ki}|Z_{1i}) = E(X_i - E(X_i|Z_{1i})|Z_{1i}) = 0$, $EI_1 = 0$ and $\sup_{z_0 \in G} |I_1| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

$I_2 = \frac{1}{nh_1^{l_{1c}}} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1}) I(Z_{1i}^d = z_{10}^d) \sum_{|j|=1}^s \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}) \frac{(Z_{1i}^c - Z_{10}^c)^j}{j!}$
$+ \frac{1}{nh_1^{l_{1c}}} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1}) I(Z_{1i}^d = z_{10}^d) \sum_{|j|=s} [\frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10*}^c, z_{10}^d) - \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10})]$
$\times \frac{(Z_{1i}^c - Z_{10}^c)^j}{j!} = I_{21} + I_{22}$.
Consider for $1 \le |k| \le s$,
$I_{211} = \frac{1}{nh_1^{l_{1c}}} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1}) I(Z_{1i}^d = z_{10}^d) \sum_{|j|=|k|} \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}) \frac{(Z_{1i}^c - Z_{10}^c)^j}{j!}$,
with A2(3) and by Lemma 1, $\sup_{z_{10} \in G_1} |I_{211} - EI_{211}| = O_p(h_1^{|k|}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.
$EI_{211} = \sum_{|j|=|k|} \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}) \int K_1(\Psi_i) \frac{(h_1 \Psi_i)^j}{j!} [f_1(z_{10})$
$+ \sum_{|m|=1}^{s-1} \frac{\partial^m}{\partial (z_{10}^c)^m} f_1(z_{10}) \frac{(h_1 \Psi)^m}{m!} + \sum_{|m|=s} \frac{\partial^m}{\partial (z_{10}^c)^m} f_1(z_{10*}^c, z_{10}^d) \frac{(h_1 \Psi)^m}{m!}] d\Psi_i$,
where z_{10*}^c is between z_{10}^c and z_{1i}^c. Since by A3, the kernel function is of order $3s_1$, and by A2(3), $\int K_1(\Psi_i) \Psi_i^{j+m} \frac{\partial^m}{\partial (z_{10}^c)^m} f_1(z_{10*}^c, z_{10}^d) d\Psi_i$
$\to \frac{\partial^m}{\partial (z_{10}^c)^m} f_1(z_{10}) \int K_1(\Psi_i) \Psi_i^{j+m} d\Psi_i < \infty$, so $EI_{211} = O(h_1^{s+k})$, and $\sup_{z_{10} \in G_1} |I_{21}| = O(h_1^{s+1}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.
$EI_{22} = h_1^s \sum_{|j|=s} \frac{1}{j!} \int K_1(\Psi_i) \Psi_i^j [\frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}^c + \lambda h_1 \Psi_i, z_{10}^d) - \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10})] f_1(z_{10}^c + h_1 \Psi_i, z_{10}^d) d\Psi_i$,
by A2(3), $\frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10})$ is uniformly continuous around $z_{10}^c \in G_1^c$,
$\le h_1^{s+1} \sum_{|j|=s} \frac{1}{j!} \int |K_1(\Psi_i) \Psi_i^j | c\lambda || \Psi_i || f_1(z_{10}^c + h_1 \Psi_i, z_{10}^d) d\Psi_i = O_p(h_1^{s+1})$ by A2(2) and A3(1).
By applying Lemma 1 with $w(Z_{1i}^c - z_{10}^c; z_0) = \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10*}^c, z_{10}^d) - \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10})$, by A2(3) $|w(Z_{1i}^c - z_{10}^c; z_0)| < \infty$, and
$|w(Z_{1i}^c - z_{10}^c; z_0) - w(Z_{1i}^c - z_{1k}^c; z_k^c, z_0^d)|$
$\le |\frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}^c + \lambda(Z_{1i}^c - z_{10}^c), z_{10}^d) - \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{1k}^c + \lambda(Z_{1i} - Z_{1k}^c), z_{10}^d)|$
$+ |\frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{1k}^c, z_{10}^d) - \frac{\partial^j}{\partial (z_{10}^c)^j} g_{1,k}(z_{10}^c, z_{10}^d)|$
$\le ||(1-\lambda)(z_{10}^c - z_{1k}^c)|| + ||z_{10}^c - z_{1k}^c|| \le (2-\lambda)||z_{10} - z_{1k}||$,
so $\sup_{z_{10} \in G_1} |I_{22} - EI_{22}| = O_p(h_1^s(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$ and $\sup_{z_{10} \in G_1} |I_{22}| = O(h_1^{s+1}) + O_p(h_2^s(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$. So in all, $\sup_{z_{10} \in G_1} |I_2| = O(h_1^{s+1}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

(d) We show $\sup_{z_{10} \in G_1} |\hat{g}_{1,j}(z_{10}) - g_{1,j}(z_{10})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})$.
$\sup_{z_{10} \in G_1} |\hat{g}_{1,j}(z_{10}) - g_{1,j}(z_{10})| = \sup_{z_{10} \in G_1} |\frac{1}{nh_1^{l_{1c}}} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1})$
$\times I(Z_{1i}^d = z_{10}^d)[X_{i,j} - g_{1,j}(z_{10})][\frac{f_1(z_{10}) - \hat{f}_1(z_{10})}{\hat{f}_1(z_{10}) f_1(z_{10})} + \frac{1}{f_1(z_{10})}]|$
By A2(2), $f_1(z_{10}) > 0$. $\inf_{z_{10} \in G_1} \hat{f}_1(z_{10}) \ge \inf_{z_{10} \in G_1} [\hat{f}_1(z_{10}) - f_1(z_{10})] + inf_{z_{10} \in G_1} f_1(z_{10}) > 0$, since in (a), $\sup_{z_{10}} |\hat{f}_1(z_{10}) - f_1(z_{10})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1^s)$, $\inf_{z_{10} \in G_1} [\hat{f}_1(z_{10}) - f_1(z_{10})] \le \inf_{z_{10} \in G_1} |\hat{f}_1(z_{10}) - f_1(z_{10})| \le \sup_{z_{10} \in G_1} |\hat{f}_1(z_{10}) - f_1(z_{10})| = o_p(1)$.
So $\sup_{z_{10} \in G_1} |\hat{g}_{1,j}(z_{10}) - g_{1,j}(z_{10})| = [O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})][O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1^s) + \frac{1}{f_1(z_{10})}] = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})$.
Furthermore, $\sup_{z_{10} \in G_1} |\hat{g}_{1,j}(z_{10}) - g_{1,j}(z_{10}) - \frac{1}{nh_1^{l_{1c}} f_1(z_{10})} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - z_{10}^c}{h_1}) I(Z_{1i}^d = z_{10}^d)[X_{i,j} - g_{1,j}(z_{10})]| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-1}) + O(h_1^{2s+1}) + O_p(h_1^s(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

(e) We can show similarly $\sup_{z_0\in G}|\frac{1}{nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-z_{10}^c}{h_2},\frac{Z_{2i}^c-z_{20}^c}{h_2})I(Z_i^d=z_0^d)[X_{i,k}-g_k(z_0)]|=$ $O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})+O(h_2^{s_1+1})$ with A2(4)-(6), A3, A4(1) and Lemma 1.

(f) We show $\sup_{z_0\in G}|\hat{g}_j(z_0)-g_j(z_0)|=O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})+O(h_2^{s_1+1})$.

$\sup_{z_0\in G}|\hat{g}_j(z_0)-g_j(z_0)|=\sup_{z_0\in G}|[\frac{1}{nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-z_{10}^c}{h_2},\frac{Z_{2i}^c-z_{20}^c}{h_2})$
$\times I(Z_i^d=z_0^d)[X_{i,j}-g_j(z_0)]][\frac{f(z_0)-\hat{f}(z_0)}{\hat{f}(z_0)f(z_0)}+\frac{1}{f(z_0)}]|$
$=[O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})+O(h_2^{s_1+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})+O_p(h_2^{s_1})+\frac{1}{f(z_0)}]$
$=O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})+O(h_2^{s_1+1})$.

Furthermore, $\sup_{z_0\in G}|\hat{g}_j(z_0)-g_j(z_0)-\frac{1}{nh_2^{l_{1c}+l_{2c}}f_z(Z_0)}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-z_{10}^c}{h_2},\frac{Z_{2i}^c-z_{20}^c}{h_2})I(Z_i^d=z_0^d)[X_{i,j}-g_j(z_0)]|=O_p((\frac{nh_1^{l_{1c}+l_{2c}}}{ln(n)})^{-1})+O(h_2^{2s_1+1})+O_p(h_2^{s_1}(\frac{nh_1^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})$.

$A_1=\frac{1}{n}\sum_t[\hat{g}_i(Z_t)-g_i(Z_t)][\hat{g}_j(Z_t)-g_j(Z_t)]$
$=O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1})+O(h_2^{2(s_1+1)})+O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})=o_p(1)$ with result in (f) and A5.
Similarly, we use results in (d) and (f) to show A_2, A_4 and A_5 are $o_p(1)$.
$A_3\le\sup_{z_0\in G}|\hat{g}_i(z_0)-g_i(z_0)|\frac{1}{n}\sum_t|g_j(Z_t)-g_{1,j}(Z_{1t})|=o_p(1)\frac{1}{n}\sum_t|g_j(Z_t)-g_{1,j}(Z_{1t})|$, since Z_t is iid, by Khinchin's theorem, $\frac{1}{n}\sum_t|g_j(Z_t)-g_{1,j}(Z_{1t})|\xrightarrow{p}E|g_j(Z_t)-g_{1,j}(Z_{1t})|$, provided $E|g_j(Z_t)-g_{1,j}(Z_{1t})|<\infty$. Since $E|g_j(Z_t)-g_{1,j}(Z_{1t})|=E|E(X_{t,j}|Z_t)-E(X_{t,j}|Z_{1t})|<2E(|X_t|)<\infty$ by A4, $A_3=o_p(1)$. Similar arguments show that A_6, A_7 and A_8 are $o_p(1)$.
By Khinchin's theorem, $A_9\xrightarrow{p}E[g_i(Z_t)-g_{1,i}(Z_{1t})][g_j(Z_t)-g_{1,j}(Z_{1t})]$, which is the $(i,j)th$ element of $EW_t'W_t$, provided $E[g_i(Z_t)-g_{1,i}(Z_{1t})][g_j(Z_t)-g_{1,j}(Z_{1t})]<\infty$.
$E[g_i(Z_t)-g_{1,i}(Z_{1t})][g_j(Z_t)-g_{1,j}(Z_{1t})]\le E|g_i(Z_t)g_i(Z_t)|+E|g_i(Z_t)g_{1,j}(Z_{1t})|+E|g_{1,i}(Z_{1t})g_j(Z_t)|+E|g_{1,i}(Z_{1t})g_{1,j}(Z_{1t})|\le\infty$ by Cauchy-Schwartz inequality and A4(1).

(2) (a) We first show
$\sup_{Z_{1t}\in G_1}|\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c-Z_{1t}^c}{h_1})I(Z_{1i}^d=Z_{1t}^d)[m(Z_{1i}-m(Z_{1t})]$
$O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O(h_1^{s+1})$, following similar arguments as in (1)(c) I_2, using A2(7), A3 and Lemma 1.

(b) We can show similarly $\sup_{Z_{1t}\in G_1}|\hat{E}(m(z_1)|Z_{1t})-m(Z_{1t})|$
$=\sup_{Z_{1t}\in G_1}|\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c-Z_{1t}^c}{h_1})I(Z_{1i}^d=Z_{1t}^d)[m(Z_{1i}-m(Z_{1t})]$
$\times[\frac{f_1(Z_{1t})-\hat{f}_1(Z_{1t})}{\hat{f}_1(Z_{1t})f_1(Z_{1t})}+\frac{1}{f_1(Z_{1t})}]|$
$=[O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O(h_1^{s+1})][O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O_p(h_1^s)+\frac{1}{f_1(Z_{1t})}]$ with A2(1)-(2), A3, Lemma 1 and (2)(a),
$=O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O(h_1^{s+1})$.
Similarly, we obtain $\sup_{Z_{1t}\in G_1}|\hat{E}(m(z_1)|Z_{1t})-m(Z_{1t})$
$-\frac{1}{f_1(Z_{1t})nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c-Z_{1t}^c}{h_1})I(Z_{1i}^d=Z_{1t}^d)[m(Z_{1i}-m(Z_{1t})]|$
$=[O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O(h_1^{s+1})][O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O_p(h_1^s)]$
$=O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-1})+O(h_1^{2s+1})+O_p(h_1^{s+1}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

(c) We show
$\sup_{Z_t\in G}|\frac{1}{nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2},\frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d=Z_t^d)[m(Z_{1i}-m(Z_{1t})]$

$O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$, following similar arguments as in (1)(c) I_2, using A2(7), A3 and Lemma 1.

(d) We can show similarly $\sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - m(Z_{1t})|$
$= \sup_{Z_t\in G}|\frac{1}{nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2}, \frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d = Z_t^d)[m(Z_{1i} - m(Z_{1t})]$
$\times[\frac{f(Z_t)-\hat{f}(Z_t)}{\hat{f}(Z_t)f(Z_t)} + \frac{1}{f(Z_t)}]|$
$= [O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_2^{s_1}) + \frac{1}{f(Z_t)}]$ with A2(4)-(5), A3, Lemma 1 and (2)(a),
$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$.
Similarly, we obtain $\sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - m(Z_{1t})$
$-\frac{1}{f(Z_t)nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2}, \frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d = Z_t^d)[m(Z_{1i} - m(Z_{1t})]|$
$= [O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_2^{s_1})]$
$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1}) + O(h_2^{2s_1+1}) + O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})$.

(e) With (2)(b) and (d), $\sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})|$
$\leq \sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - m(Z_{1t})| + \sup_{Z_t\in G}|m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t})|$
$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})$.
Also denote $I_3 = \frac{1}{f_1(Z_{1t})nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c-Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)[m(Z_{1i}) - m(Z_{1t})]$,
$I_4 = \frac{1}{f(Z_t)nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2}, \frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d = Z_t^d)[m(Z_{1i} - m(Z_{1t})]$,
$\sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + I_3 - I_4|$
$\leq \sup_{Z_t\in G}|\hat{E}(m(z_1)|Z_t) - m(Z_{1t}) - I_4| + \sup_{Z_{1t}\in G_1}|\hat{E}(m(z_1)|Z_{1t}) - m(Z_{1t}) - I_3|$
$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1}) + O(h_2^{2s_1+1}) + O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-1})$
$+O(h_1^{2s+1}) + O_p(h_1^{s+1}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

(f) $\sup_{Z_t\in G}|\hat{E}(\epsilon|Z_t)|$
$= \sup_{Z_t\in G}|\frac{1}{nh_2^{l_{1c}+l_{2c}}}\sum_{i=1}^n K_2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2}, \frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d = Z_t^d)\epsilon_i[\frac{f(Z_t)-\hat{f}(Z_t)}{\hat{f}(Z_t)f(Z_t)} + \frac{1}{f(Z_t)}]|$
$= O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}})$ with Lemma 1, A1(2), A2(4),(5), A3 and A4(2).

Similarly, we have $\sup_{Z_{1t}\in G_1}|\hat{E}(\epsilon|Z_{1t})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

$C_{1k} \leq \sup_{Z_t\in G}|\hat{g}_k(Z_t) - g_k(Z_t)|[\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})]$
$= [O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$
$+O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})]$ by (1)(f) and (2)(e).

$C_{2k} \leq \sup_{Z_{1t}\in G_1}|\hat{g}_{1,k}(Z_{1t}) - g_{1,k}(Z_{1t})|[\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})]$
$= [O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})][O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$
$+O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})]$ by (1)(d) and (2)(e).

$C_{3k} = \frac{1}{n}\sum_t[g_k(Z_t) - g_{1,K}(Z_t)][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + I_3 - I_4]$
$-\frac{1}{n}\sum_t[g_k(Z_t) - g_{1,K}(Z_t)]I_3 + \frac{1}{n}\sum_t[g_k(Z_t) - g_{1,K}(Z_t)]I_4$
$= C_{31k} - C_{32k} + C_{33k}$

$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1}) + O(h_2^{2s_1+1}) + O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-1})$
$\quad +O(h_1^{2s+1}) + O_p(h_1^{s+1}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O((n^2 h_1^{l_{1c}-2})^{-\frac{1}{2}}) + O(h_1^{s+1})$
$\quad +O((n^2 h_2^{l_{1c}+l_{2c}-2})^{-\frac{1}{2}}) + O(h_2^{s_1+1})$.
As in (1) item A3, $\frac{1}{n}\sum_t |g_k(Z_t) - g_{1,K}(Z_t)| = O_p(1)$, so we use (2)(e) to obtain
$C_{31k} \leq \frac{1}{n}\sum_t g_k(Z_t) - g_{1,K}(Z_t)| \sup_{Z_t \in G} |\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + I_3 - I_4|$
$= O_p(h_2(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-1}) + O(h_2^{2s_1+1}) + O_p(h_2^{s_1+1}(\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1(\frac{nh_1^{l_{1c}}}{ln(n)})^{-1})$
$\quad +O(h_1^{2s+1}) + O_p(h_1^{s+1}(\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$.

$$C_{32k} = \frac{1}{n^2 h_1^{l_{1c}}}\sum_{t\neq i}\sum_i \underbrace{\frac{g_k(Z_t) - g_{1,K}(Z_{1t})}{f_1(Z_{1t})} K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_1}) I(Z_{1i}^d = Z_{1t}^d)[m(Z_{1i}) - m(Z_{1t})]}_{\Psi(Z_i, Z_t)}$$

$= \frac{1}{2n^2 h_1^{l_{1c}}}\sum_{t\neq i}\sum_i [\Psi(Z_i, Z_t) + \Psi_(Z_t, Z_i)] = \frac{1}{2n^2 h_1^{l_{1c}}}\sum_{t\neq i}\sum_i \phi(Z_i, Z_t)$.
Let's define $\hat{E}\phi(Z_i, Z_t) = \int \phi(Z_i, Z_t) f(Z_i) dZ_i + \int \phi(Z_i, Z_t) f(Z_t) dZ_t - E\phi(Z_i, Z_t)$, and $\Phi(Z_i, Z_t) = \phi(Z_i, Z_t) - \hat{E}\phi(Z_i, Z_t)$. Note $\Phi(Z_i, Z_t)$ is symmetric and has conditional mean zero by construction. $\hat{E}\phi(Z_i, Z_t)$ is the U-statistics projection.
$C_{32k} = \frac{1}{n^2 h_1^{l_{1c}}}\sum_{t<i}\sum_i (\Phi(Z_i, Z_t) + \hat{E}\phi(Z_i, Z_t))$. By c_r inequality
$EC_{32k}^2 \leq \frac{c}{n^4 h_1^{2l_c}}[E(\sum_{t<i}\sum_i \Phi(Z_i, Z_t))^2 + E(\sum_{t<i}\sum_i \hat{E}\phi(Z_i, Z_t))^2]$
$\quad = \frac{c}{n^4 h_1^{2l_{1c}}}(C_{32ak} + C_{32bk})$.
If t, i, t', i' are different, $C_{32ak} = \sum_t \sum_i \sum_{t'} \sum_{i', t<i, t'<i'} E\Phi(Z_i, Z_t)\Phi(Z_{i'}, Z_{t'}) = 0$, since the conditional mean of $\Phi(Z_i, Z_t)$ is zero.
If only three of the four indices in the sum are different, for example, t, i, t' are different, $C_{32ak} = \sum_t \sum_i \sum_{t', t<i, t'<t} E\Phi(Z_i, Z_t)\Phi(Z_t, Z_{t'}) = 0$.
If only two of the four indices in the sum are different,
$C_{32ak} = \sum_t \sum_{i, t<i} E\Phi^2(Z_i, Z_t) = \frac{n(n-1)}{2} E\Phi^2(Z_i, Z_t)$.
$E\Phi^2(Z_i, Z_t) = E[\phi(Z_i, Z_t) - \int \phi(Z_i, Z_t) f(Z_i) dZ_i - \int \phi(Z_i, Z_t) f(Z_t) dZ_t + E\phi(Z_i, Z_t)]^2$
$\leq c_r[E\phi^2(Z_i, Z_t) + 3\int\int \phi^2(Z_i, Z_t) f(Z_i) f(Z_t) dZ_i dZ_t]$, by c_r inequality and Cauchy-Schwartz inequality,
$E\phi^2(Z_i, Z_t) \leq 2[E\Psi^2(Z_i, Z_t) + E\Psi^2(Z_t, Z_i)]$ by c_r inequality again,
$E\Psi^2(Z_i, Z_t) = E\frac{(g_k(Z_t) - g_{1,K}(Z_t))^2}{f_1^2(Z_{1t})} K_1^2(\frac{Z_{1i}^c - Z_{1t}^c}{h_1}) I(Z_{1i}^d = Z_{1t}^d)[m(Z_{1i}) - m(Z_{1t})]^2$
$\leq h_1^{l_{1c}} \sum_{Z_t^d} \int \frac{(g_k(Z_t) - g_{1,K}(Z_t))^2}{f_1^2(Z_{1t})} K_1^2(\Psi_i)[\frac{\partial m(Z_{1t*}^c, Z_{1t}^d)}{\partial Z_{1t}^c} h_1 \Psi_i]^2$
$\quad \times f_1(Z_{1t}^c + h\Psi_i, Z_{1t}^d) f(Z_t) d\Psi_i dZ_{1t}^c dZ_{2t}^c = O(h_1^{l_{1c}+2})$, by A2(1), (2),(7), A3 and A4(1), for Z_{1t*}^c between Z_{1t}^c and Z_{1i}^c. So $E\phi^2(Z_i, Z_t) = O(h_1^{l_{1c}+2})$.
Since $Z_i's$ are IID by A1(1), $\int\int \phi^2(Z_i, Z_t) f(Z_i) f(Z_t) dZ_i dZ_t = E\phi^2(Z_i, Z_t) = O(h_1^{l_{1c}+2})$.
In all, we have $C_{32ak} = O(n^2 h_1^{l_{1c}+2})$.

$$\begin{aligned}
C_{32bk} &= 4E[\textstyle\sum_t \sum_{i,t<i} \hat{E}\phi(Z_i, Z_t)]^2 = 4[\sum_t \sum_{i,t<i} E(\hat{E}\phi(Z_i, Z_t))^2 \\
&\quad + \textstyle\sum_t \sum_{i,t<i} \sum_{t'} \sum_{i', t'<i', (t,i)\neq(t',i')} E(\hat{E}\phi(Z_i, Z_t))(\hat{E}\phi(Z_{i'}, Z_{t'}))] \\
&= 4[C_{32b1k} + C_{32b2k}] \\
C_{32b1k} &= \textstyle\sum_t \sum_{i,t<i} E[\int \phi(Z_i, Z_t) f(Z_i) dZ_i + \int \phi(Z_i, Z_t) f(Z_t) dZ_t - E\phi(Z_i, Z_t)]^2 \\
&\leq cn^2[E(\textstyle\int \phi(Z_i, Z_t) f(Z_i) dZ_i)^2 + E(\int \phi(Z_i, Z_t) f(Z_t) dZ_t)^2 + E^2\phi(Z_i, Z_t)] \;. \\
&= cn^2[2E(\textstyle\int \phi(Z_i, Z_t) f(Z_i) dZ_i)^2 + E^2\phi(Z_i, Z_t)]
\end{aligned}$$

$\int \phi(Z_i, Z_t) f(Z_i) dZ_i = \int \Psi(Z_i, Z_t) f(Z_i) dZ_i + \int \Psi(Z_t, Z_i) f(Z_i) dZ_i$
$\quad = \frac{g_k(Z_t) - g_{1,K}(Z_t)}{f_1(Z_{1t})} O(h_1^{l_{1c}+s+1}) + O(h_1^{l_{1c}+s+1})$. it follows from results below.
$\int \Psi(Z_i, Z_t) f(Z_i) dZ_i = h_1^{l_{1c}} \frac{g_k(Z_t) - g_{1,K}(Z_t)}{f_1(Z_{1t})}[\int K_1(\Psi_i)(\sum_{|j|=1}^s \frac{\partial^j}{\partial(Z_{1t}^c)^j} m(Z_{1t}) \frac{(h_1\Psi_i)^j}{j!})$

$\times(f_1(Z_{1t})+\sum_{|m|=1}^{s-1}\frac{\partial^m}{\partial(Z_{1t}^c)^m}f_1(Z_{1t})\frac{(h_1\Psi_i)^m}{m!}+\sum_{|m|=s}\frac{\partial^m}{\partial(Z_{1t}^c)^m}f_1(Z_{1t*}^c,Z_{1t}^d)\frac{(h_1\Psi_i)^m}{m!})d\Psi$

$+\int K_1(\Psi_i)\sum_{|j|=s}(\frac{\partial^j}{\partial(Z_{1t}^c)^j}m(Z_{1t*}^c,Z_{1t}^d)-\frac{\partial^j}{\partial(Z_{1t}^c)^j}m(Z_{1t}))\frac{(h_1\Psi_i)^j}{j!}f_1(Z_{1t}^c+h\Psi_i,Z_{1t}^d)d\Psi_i]$

$=\frac{g_k(Z_t)-g_{1,K}(Z_t)}{f_1(Z_{1t})}O(h_1^{l_{1c}+s+1})$ by A2(1), (7) and A3 using similar arguments as in (1)(c).

$\int\Psi(Z_t,Z_i)f(Z_i)dZ_i$

$=h_1^{l_{1c}}\sum_{Z_{2i}^d}\int[\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}+\sum_{|j|=1}^{s}\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}\frac{(-h_1\Psi_i)^j}{j!}$

$+\sum_{|j|=s}(\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t*}^c,Z_{1t}^d,Z_{2i})-g_{1,K}(Z_{1t*}^c,Z_{1t}^d)}{f_1(Z_{1t*}^c,Z_{1t}^d)}-\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})})\frac{(-h_1\Psi_i)^j}{j!}]$

$\times[\sum_{|m|=1}^{s}\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t})\frac{(-h_1\Psi_i)^m}{m!}+\sum_{|m|=s}(\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t*}^c,Z_{1t}^d)-\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t})\frac{(-h_1\Psi_i)^m}{m!}]$

$\times K_1(\Psi_i)[f(Z_{1t},Z_{2i})+\sum_{|l|=1}^{s-1}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}$

$+\sum_{|l|=s}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t*}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}]d\Psi_i dZ_{2i}^c=O(h_1^{l_{1c}+s+1})$.

Above results are obtained using

(i) $\sum_{Z_{2i}^d}\int\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}[\sum_{|m|=1}^{s}\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t})\frac{(-h_1\Psi_i)^m}{m!}]$

$\times K_1(\Psi_i)[f(Z_{1t},Z_{2i})+\sum_{|l|=1}^{s-1}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}$

$+\sum_{|l|=s}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t*}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}]d\Psi_i dZ_{2i}^c=O(h_1^{s+1})$ using A2(2),(4),(5),(7), A3 and A4(1).

(ii) $\sum_{Z_{2i}^d}\int\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}[\sum_{|m|=s}(\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t*}^c,Z_{1t}^d)-\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t}))\frac{(-h_1\Psi_i)^m}{m!})]$

$\times K_1(\Psi_i)f(Z_{1t}^c-h_1\Psi_i,Z_{1t}^d,Z_{2i})d\Psi_i dZ_{2i}^c=O(h_1^{s+1})$ by A2(2), (4), (5),(7),A3, and A4(1).

(iii) $\sum_{Z_{2i}^d}\int\sum_{|j|=1}^{s}\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}\frac{(-h_1\Psi_i)^j}{j!}[\sum_{|m|=1}^{s}\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t})\frac{(-h_1\Psi_i)^m}{m!}]$

$\times K_1(\Psi_i)[f(Z_{1t},Z_{2i})+\sum_{|l|=1}^{s-1}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}$

$+\sum_{|l|=s}\frac{\partial^l}{\partial(Z_{1t}^c)^l}f(Z_{1t*}^c,Z_{1t}^d,Z_{2i})\frac{(-h_1\Psi_i)^l}{l!}]d\Psi_i dZ_{2i}^c=O(h_1^{s+2})$ using A2, A3 and A4(1).

(iv) $\sum_{Z_{2i}^d}\int\sum_{|j|=1}^{s}\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})}\frac{(-h_1\Psi_i)^j}{j!}$

$\times[\sum_{|m|=s}(\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t*}^c,Z_{1t}^d)-\frac{\partial^m}{\partial(Z_{1t}^c)^m}m(Z_{1t}))\frac{(-h_1\Psi_i)^m}{m!})]\times K_1(\Psi_i)f(Z_{1t}^c-h_1\Psi_i,Z_{1t}^d,Z_{2i})d\Psi_i dZ_{2i}^c$

$=O(h_1^{s+2})$.

(v) $\sum_{Z_{2i}^d}\int\sum_{|j|=s}(\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t*}^c,Z_{1t}^d,Z_{2i})-g_{1,K}(Z_{1t*}^c,Z_{1t}^d)}{f_1(Z_{1t*}^c,Z_{1t}^d)}-\frac{\partial^j}{\partial(Z_{1t}^c)^j}\frac{g_k(Z_{1t},Z_{2i})-g_{1,K}(Z_{1t})}{f_1(Z_{1t})})\frac{(-h_1\Psi_i)^j}{j!}]$

$\times(m(Z_{1t})-m(Z_{1t}^c-h_1\Psi_i,Z_{qt}^d))K_1(\Psi_i)f(Z_{1t}^c-h_1\Psi_i,Z_{1t}^d,Z_{2i})d\Psi_i dZ_{2i}^c=O(h_1^{s+1})$.

$E(\int\phi(Z_i,Z_t)f(Z_i)dZ_i)^2\leq 2[E(\int\Psi(Z_i,Z_t)f(Z_i)dZ_i)^2+E(\int\Psi(Z_t,Z_i)f(Z_i)dZ_i)^2]$

$=2E(\frac{g_k(Z_t)-g_{1,K}(Z_t)}{f_1(Z_{1t})})^2O(h_1^{2(l_{1c}+s+1)})+O(h_1^{2(l_{1c}+s+1)})=O(h_1^{2(l_{1c}+s+1)})$ by A4(1).

$E^2\phi(Z_i,Z_t)=[\int\int\phi(Z_i,Z_t)f(Z_i)dZ_if(Z_t)dZ_t]^2=O(h_1^{2(l_{1c}+s+1)})$.

So $C_{32b1k}=O(n^2h_1^{2(l_{1c}+s+1)})$.

If t,i,t',i' are all different,

$$\begin{aligned}
C_{32b2k}\leq\ & n^4E[\int\phi(Z_i,Z_t)f(Z_i)dZ_i+\int\phi(Z_t,Z_i)f(Z_i)dZ_i-E(\phi(Z_i,Z_t))]\\
&\times[\int\phi(Z_{i'},Z_{t'})f(Z_{i'})dZ_{i'}+\int\phi(Z_{t'},Z_{i'})f(Z_{i'})dZ_{i'}-E(\phi(Z_{i'},Z_{t'}))]\\
=\ & n^4[E(\int\phi(Z_i,Z_t)f(Z_i)dZ_i)(\int\phi(Z_{i'},Z_{t'})f(Z_{i'})dZ_{i'})\\
&+2E(\int\phi(Z_i,Z_t)f(Z_i)dZ_i)(\int\phi(Z_{i'},Z_{t'})f(Z_{i'})dZ_{i'})\\
&+E(\int\phi(Z_t,Z_i)f(Z_i)dZ_i)(\int\phi(Z_{t'},Z_{i'})f(Z_{i'})dZ_{i'})-3E^2(\phi(Z_{i'},Z_{t'}))]\\
\leq\ & n^4\{[E(\int\phi(Z_i,Z_t)f(Z_i)dZ_i)^2E(\int\phi(Z_{i'},Z_{t'})f(Z_{i'})dZ_{i'})^2]^{\frac{1}{2}}\\
&+2[E(\int\phi(Z_i,Z_t)f(Z_i)dZ_i)^2E(\int\phi(Z_{i'},Z_{t'})f(Z_{i'})dZ_{i'})^2]^{\frac{1}{2}}\\
&+[E(\int\phi(Z_t,Z_i)f(Z_i)dZ_i)^2E(\int\phi(Z_{t'},Z_{i'})f(Z_{i'})dZ_{i'})^2]^{\frac{1}{2}}-3E^2(\phi(Z_{i'},Z_{t'}))\}\\
=\ & O(n^4h_1^{2(l_{1c}+s+1)})\text{ by Cauchy-Schwartz inequality.}
\end{aligned}$$

If only three indices, say, t,i,t' are different, $C_{32b2k}=O(n^3h_1^{2(l_{1c}+s+1)})$. If there are only two distinct indices, $t<i$, then $C_{32b2k}=O(n^2h_1^{2(l_{1c}+s+1)})$. So in all we have $C_{32b2k}=O(n^4h_1^{2(l_{1c}+s+1)})$ and

$C_{32bk} = O(n^4 h_1^{2(l_{1c}+s+1)})$.
Since $E^2 C_{32k} \leq \frac{c}{4n^4 h_1^{2l_c}}(C_{32ak} + C_{32bk}) = O((n^2 h_1^{l_{1c}-2})^{-1}) + O(h_1^{2(s+1)})$, so
$\sqrt{n} C_{32k} = O_p((nh_1^{l_{1c}-2})^{-\frac{1}{2}}) + O_p((nh_1^{2(s+1)})^{\frac{1}{2}}) = o_p(1)$.

$C_{33k} = \frac{1}{n^2 h_2^{l_{1c}+l_{2c}}} \sum_{t \neq i} \sum_i \frac{g_k(Z_t) - g_{1,K}(Z_{1t})}{f(Z_t)} K_2(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}, \frac{Z_{2i}^c - Z_{2t}^c}{h_2}) I(Z_i^d = Z_t^d)[m(Z_{1i}) - m(Z_{1t})]$. With assumption A2 that $g_{1,k}(Z_{1t})$, $f(Z_t)$, $g_k(Z_t)$ and $m(Z_{1t})$ are s_1 times continuously differentiable with the $s_1 th$ order derivative uniformly continuous, with uniformly bounded derivatives, we similarly obtain
$E^2 C_{33k} = O((n^2 h_2^{l_{1c}+l_{2c}-2})^{-1}) + O(h_2^{2(s_1+1)})$, so
$\sqrt{n} C_{33k} = O_p((nh_2^{l_{1c}+l_{2c}-2})^{-\frac{1}{2}}) + O_p((nh_2^{2(s_1+1)})^{\frac{1}{2}}) = o_p(1)$.

$C_{4k} \leq \sup_{Z_t \in G} |\hat{g}_k(Z_t) - g_k(Z_t)||\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})|$
$= [O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_2^{s_1+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})]$ with (1)(f) and (2)(f).

$C_{5k} \leq \sup_{Z_t \in G} |\hat{g}_{1,k}(Z_{1t}) - g_{1,k}(Z_{1t})||\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})|$
$= [O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1})][O_p((\frac{nh_2^{l_{1c}+l_{2c}}}{ln(n)})^{-\frac{1}{2}}) + O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})]$ with (1)(d) and (2)(f).

$C_{6k} = \frac{1}{n} \sum_t [g_k(Z_t) - g_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})]$
$= \frac{1}{n} \sum_t [g_k(Z_t) - g_{1,k}(Z_{1t})]\{[\frac{f(Z_t) - \hat{f}(Z_t)}{\hat{f}(Z_t) f(Z_t)} + \frac{1}{f(Z_t)}]$
$\times [\frac{1}{nh_2^{l_{1c}+l_{2c}}} \sum_{i=1}^n K_2(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}, \frac{Z_{2i}^c - z_{2t}^c}{h_2}) I(Z_i^d = Z_t^d)\epsilon_i]$
$-[\frac{f_1(Z_{1t}) - \hat{f}_1(Z_{1t})}{\hat{f}_1(Z_{1t}) f_1(Z_{1t})} + \frac{1}{f_1(Z_{1t})}][\frac{1}{nh_1^{l_{1c}}} \sum_{i=1}^n K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}) I(Z_{1i}^d = Z_{1t}^d)\epsilon_i]\}$
$= \frac{1}{n^2} \sum_t \sum_i [g_k(Z_t) - g_{1,k}(Z_{1t})]\{\frac{1}{f(Z_t)} \frac{1}{h_2^{l_{1c}+l_{2c}}} K_2(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}, \frac{Z_{2i}^c - z_{2t}^c}{h_2}) I(Z_i^d = Z_t^d)\epsilon_i$
$-\frac{1}{f_1(Z_{1t})} \frac{1}{h_1^{l_{1c}}} K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}) I(Z_{1i}^d = Z_{1t}^d)\epsilon_i\}\{1 + o_p(1)\}$ where $S_i = (Z_i, \epsilon_i)$, with (1)(a), (b), and A2(2) and (5),
$= \frac{1}{n^2} \sum_t \sum_i \Psi(S_i, S_t)\{1 + o_p(1)\} = \frac{1}{2n^2} \sum_t \sum_i (\Psi(S_i, S_t) + \Psi(S_t, S_i))\{1 + o_p(1)\}$
$= \frac{1}{2n^2} \sum_t \sum_i \phi(S_i, S_t)\{1 + o_p(1)\}$, where $\phi(S_i, S_t)$ is symmetric,
$= [\frac{1}{2n^2} \sum_t \phi(S_t, S_t) + \frac{1}{n^2} \sum_t \sum_{i,t<i} \phi(S_i, S_t)]\{1 + o_p(1)\}$
$= [C_{61k} + C_{62k}]\{1 + o_p(1)\}$.

$E\phi^2(S_t, S_t) = E([g_k(Z_t) - g_{1,k}(Z_{1t})]^2 (\frac{c\epsilon_t}{h_2^{l_{1c}+l_{2c}} f(Z_t)} - \frac{c\epsilon_t}{h_1^{l_{1c}} f_1(Z_{1t})})^2$
$= O(h_2^{-2(l_{1c}+l_{2c})}) + O(h_1^{-2l_{1c}})$, by A4(1) and (3).

So $C_{61k} = O_p((nh_2^{l_{1c}+l_{2c}})^{-1}) + O_p((nh_1^{l_{1c}})^{-1})$.

$$\begin{aligned} C_{62k} &= \frac{1}{n^2} \sum_t \sum_{i,t<i} \underbrace{[\phi(S_i, S_t) - \hat{E}\phi(S_i, S_t)]}_{\Phi(S_i, S_t)} + \frac{1}{n^2} \sum_t \sum_{i,t<i} \hat{E}\phi(S_i, S_t) \\ &= C_{621k} + C_{622k} \end{aligned}$$

where $\hat{E}\phi(S_i, S_t) = E(\phi(S_i, S_t)|S_t) + E(\phi(S_i, S_t)|S_i) - E\phi(S_i, S_t) = E(\phi(S_i, S_t)|S_t) + E(\phi(S_i, S_t)|S_i)$ since $E\phi(S_i, S_t) = 0$.
$EC_{621k}^2 = \frac{1}{n^4} E(\sum_t \sum_{i,t<i} \Phi(S_i, S_t))^2 = \frac{1}{n^4} \sum_t \sum_{i,t<i} \sum_{t'} \sum_{i',t'<i'} E\Phi(S_i, S_t)\Phi(S_{i'}, S_{t'})$.
If t, i, t', i' are all different, $EC_{621k}^2 = 0$.
If only three of the four indices in the sum are different, $EC_{621k}^2 = 0$.
If only two of the four indices in the sum are different,
$EC_{621k}^2 = \frac{1}{n^4} \sum_t \sum_{i,t<i} E\Phi^2(S_i, S_t) = \frac{1}{n^4} \frac{n(n-1)}{2} E\Phi^2(S_i, S_t)$.

Since $E\Phi^2(S_i,S_t) = E(\phi(S_i,S_t) - E(\phi(S_i,S_t)|S_t) - E(\phi(S_i,S_t)|S_i))^2 \le cE\phi^2(S_i,S_t)$,
$E\phi^2(S_i,S_t) \le c[E\Psi^2(S_i,S_t) + E\Psi^2(S_t,S_i)] = 2cE\Psi^2(S_i,S_t)$.

$$\begin{aligned} E\Psi^2(S_i,S_t) &\le c[E\frac{(g_k(Z_t)-g_{1,k}(Z_{1t}))^2}{h_2^{2(l_{1c}+l_{2c})}f^2(Z_t)}K_2^2(\frac{Z_{1i}^c-Z_{1t}^c}{h_2},\frac{Z_{2i}^c-z_{2t}^c}{h_2})I(Z_i^d=Z_t^d)\epsilon_i^2 \\ &+E\frac{(g_k(Z_t)-g_{1,k}(Z_{1t}))^2}{h_1^{2l_{1c}}f_1^2(Z_{1t})}K_1^2(\frac{Z_{1i}^c-Z_{1t}^c}{h_1})I(Z_{1i}^d=Z_t^d)\epsilon_i^2] \end{aligned}$$

by A4(3), $E(\epsilon_i^2|Z_t,Z_i)<\infty$, and also by A4(1), A3, A2(2) and (5),
$E\phi^2(S_i,S_t) = O(h_2^{-(l_{1c}+l_{2c})}) + O(h_1^{-l_{1c}})$, so $EC_{621k}^2 = O(n^{-2}(h_2^{-(l_{1c}+l_{2c})}+h_1^{-l_{1c}}))$ and $C_{621k} = O(n^{-1}(h_2^{-\frac{(l_{1c}+l_{2c})}{2}}+h_1^{-\frac{l_{1c}}{2}}))$.

$\sqrt{n}C_{622k} = \sqrt{n}\frac{1}{n^2}\sum_t\sum_{i,t<i}[E(\phi(S_i,S_t)|S_t)+E(\phi(S_i,S_t)|S_i)]$
$= \sqrt{n}\frac{1}{n^2}\sum_t\sum_{i,t\ne i}E(\phi(S_i,S_t)|S_t) = \frac{n-1}{n}\sum_{t=1}^n\frac{1}{\sqrt{n}}S_{tn} = \frac{n-1}{n}\sum_{t=1}^n\tilde{S}_{tn}$, where

$$\begin{aligned} S_{tn} &= \epsilon_t[\int\frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}f_z(Z_i)}K_2(\frac{Z_{1t}^c-Z_{1i}^c}{h_2},\frac{Z_{2t}^c-z_{2i}^c}{h_2})I(Z_t^d=Z_i^d)f_z(Z_i)dZ_i \\ &-\int\frac{(g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}f_1(Z_{1i})}K_1(\frac{Z_{1t}^c-Z_{1i}^c}{h_1})I(Z_{1t}^d=Z_{1i}^d)f_z(Z_i)dZ_i]. \end{aligned}$$

Note that $\tilde{S}_{tn}$ forms am independent triangular array, $E\tilde{S}_{tn}=0$, and
$\sum_{t=1}^n E\tilde{S}_{tn}^2 = V(S_{tn}) = E\sigma^2(Z_t)[g_k(Z_t)-E_{Z_{2i}|Z_{1t}}(g_k(Z_{1t},Z_{2i}))]^2+o(1)$.
By Liapounov's Central Limit Theorem, provided $\lim_{n\to\infty}\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta}=0$ for some $\delta>0$, we have $\sum_t\tilde{S}_{tn}\xrightarrow{d}N(0,E\sigma^2(Z_t)[g_k(Z_t)-E_{Z_{2i}|Z_{1t}}(g_k(Z_{1t},Z_{2i}))]^2)$.
$\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta} = (\sum_t E\tilde{S}_{tn}^2)^{-1-\frac{\delta}{2}}n^{-\frac{\delta}{2}}E|S_{tn}|^{2+\delta}$,
$E|S_{tn}|^{2+\delta}$
$\le c\{E[E(|\epsilon_t|^{2+\delta}|Z_t)|\int\frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}f_z(Z_i)}K_2(\frac{Z_{1t}^c-Z_{1i}^c}{h_2},\frac{Z_{2t}^c-z_{2i}^c}{h_2})I(Z_t^d=Z_i^d)f_z(Z_i)dZ_i|^{2+\delta}]$
$+E[E(|\epsilon_t|^{2+\delta}|Z_t)|\int\frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}f_1(Z_{1i})}K_1(\frac{Z_{1t}^c-Z_{1i}^c}{h_1})I(Z_{1t}^d=Z_{1i}^d)f_z(Z_i)dZ_i|^{2+\delta}]\}$
Since $E(|\epsilon_t|^{2+\delta}|Z_t)<\infty$ by A4(3), and
$E|\int\frac{(g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}f_z(Z_i)}K_2(\frac{Z_{1t}^c-Z_{1i}^c}{h_2},\frac{Z_{2t}^c-z_{2i}^c}{h_2})I(Z_t^d=Z_i^d)f_z(Z_i)dZ_i|^{2+\delta}$
$\to E|g_k(Z_t)-g_{1,k}(Z_{1t})|^{2+\delta}\le 2EX_{t,k}^{2+\delta}$, and similarly,
$E|\int\frac{(g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}f_1(Z_{1i})}K_1(\frac{Z_{1t}^c-Z_{1i}^c}{h_1})I(Z_{1t}^d=Z_{1i}^d)f_z(Z_i)dZ_i|^{2+\delta}$
$\to E|E_{Z_{2i}|Z_{1t}}(g_k(Z_{1t},Z_{2i})-g_{1,k}(Z_{1t})|^{2+\delta}=0$,
So in all we have $E|S_{tn}|^{2+\delta}<\infty$ and $\lim_{n\to\infty}\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta}=0$.
Finally with Cramer-Rao device, we obtain

$$\sqrt{n}C_6\xrightarrow{d}N(0,\Phi_0).$$

Theorem 3: *Proof.*
Let's define the infeasible estimator
$\tilde{\beta}^I = (\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1}\hat{W}'\Omega^{-1}(\vec{Z}_1)(\hat{E}(Y|\vec{Z})-\hat{E}(Y|\vec{Z}_1))$, where the true $\sigma^2(Z_{1t})$ is known in $\Omega^{-1}(\vec{Z}_1)$.
In the following we show
(1) $\sqrt{n}(\tilde{\beta}^I-\beta)\xrightarrow{d}N(0,(E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1})$.
(2) $\sup_{Z_{1t}\in G_1}|\hat{\sigma}^2(Z_{1t})-\sigma^2(Z_{1t})| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O_p(h_1^{s+1})+o_p(n^{-\frac{1}{2}})$.
Result (2) might be of use by itself. Here repeated use of (2) enables us to obtain
(3) $\sqrt{n}(\tilde{\beta}^I-\tilde{\beta}^H)=o_p(1)$.
The conclusion of Theorem 3 follows from (1) and (3).

$$
\begin{aligned}
(1)\, \tilde{\beta}^I - \beta \quad &= [(\tfrac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1} - (E\tfrac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1} + (E\tfrac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1}] \\
&\quad \underbrace{\times \frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)(\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1))}_{C}
\end{aligned}
$$

(a) The $(i,j)th$ element of $\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1}$ is

$$
\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{t=1}^n \tfrac{1}{\sigma^2(Z_{1t})}\hat{W}_{t,i}\hat{W}_{t,j} \\
= \ & \tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[\hat{g}_i(Z_t) - g_i(Z_t)][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[\hat{g}_i(Z_t) - g_i(Z_t)][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[\hat{g}_i(Z_t) - g_i(Z_t)][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_{1,i}(Z_{1t}) - \hat{g}_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_i(Z_t) - g_{1,i}(Z_{1t})][\hat{g}_j(Z_t) - g_j(Z_t)] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_i(Z_t) - g_{1,i}(Z_{1t})][g_{1,j}(Z_{1t}) - \hat{g}_{1,j}(Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_i(Z_t) - g_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] \\
= \ & A_1 + A_2 + \cdots + A_9
\end{aligned}
$$

Since Z_{1t} is iid, $\frac{1}{n}\sum_t \frac{1}{\sigma^2(Z_{1t})} \xrightarrow{p} E\frac{1}{\sigma^2(Z_{1t})} < \infty$ by A6(1), we follow the proof of Theorem 1 to obtain $A_i = o_p(1)$, $i = 1, \cdots, 8$,

$$A_9 - E\frac{1}{\sigma^2(Z_{1t})}[g_i(Z_t) - g_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] = o_p(1),$$

provided $E\frac{1}{\sigma^2(Z_{1t})}[g_i(Z_t) - g_{1,i}(Z_{1t})][g_j(Z_t) - g_{1,j}(Z_{1t})] < \infty$, which is true given A6(1) and A4(1). So together we have $\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W} - E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t = o_p(1)$. By $A6(2)$ and Slutsky' Theorem, $(\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1} - (E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1} = o_p(1)$.

(b) Similarly, for $k = 1, 2, \cdots, K$, the kth element of C is

$$
\begin{aligned}
C_k = \ & \tfrac{1}{n}\textstyle\sum_{t=1}^n \tfrac{1}{\sigma^2(Z_{1t})}\hat{W}_{t,k}(\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t}) + \hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})) \\
= \ & \tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[\hat{g}_k(Z_t) - g_k(Z_t)][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_{1,k}(Z_{1t}) - \hat{g}_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_k(Z_t) - g_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t) - \hat{E}(m(z_1)|Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[\hat{g}_k(Z_t) - g_k(Z_t)][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_{1,k}(Z_{1t}) - \hat{g}_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
& +\tfrac{1}{n}\textstyle\sum_t \tfrac{1}{\sigma^2(Z_{1t})}[g_k(Z_t) - g_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t) - \hat{E}(\epsilon|Z_{1t})] \\
= \ & C_{1k} + C_{2k} + \cdots + C_{6k}
\end{aligned}
$$

Since $\frac{1}{n}\sum_t \frac{1}{\sigma^2(Z_{1t})} \xrightarrow{p} E\frac{1}{\sigma^2(Z_{1t})} < \infty$, we follow proof of Theorem 1 to obtain $C_{ik} = o_p(n^{-\frac{1}{2}})$ for $i = 1, 2, 3, 4, 5$ with the additional assumption A6(1).

$C_{6k} = \frac{1}{n^2}\sum_t\sum_i \frac{[g_k(Z_t) - g_{1,k}(Z_{1t})]}{\sigma^2(Z_{1t})}\{\frac{1}{f(Z_t)}\frac{1}{h_2^{l_{1c}+l_{2c}}}K_2(\frac{Z_{1i}^c - Z_{1t}^c}{h_2}, \frac{Z_{2i}^c - z_{2t}^c}{h_2})I(Z_i^d = Z_t^d)\epsilon_i$

$-\frac{1}{f_1(Z_{1t})}\frac{1}{h_1^{l_{1c}}}K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_2})I(Z_{1i}^d = Z_{1t}^d)\epsilon_i\}\{1 + o_p(1)\}$

where $S_i = (Z_i, \epsilon_i)$, with (1)(a), (b), and A2(2) and (5),

$= \frac{1}{n^2}\sum_t\sum_i \Psi(S_i, S_t)\{1 + o_p(1)\} = \frac{1}{2n^2}\sum_t\sum_i \phi(S_i, S_t)\{1 + o_p(1)\}$,

where $\phi(S_i, S_t) = \Psi(S_i, S_t) + \Psi(S_t, S_i)$ is symmetric,

$= [\frac{1}{2n^2}\sum_t \phi(S_t, S_t) + \frac{1}{n^2}\sum_t\sum_{i,t<i}\phi(S_i, S_t)]\{1 + o_p(1)\}$

$= [C_{61k} + C_{62k}]\{1 + o_p(1)\}$.

$C_{61k} = O_p((nh_2^{l_{1c}+l_{2c}})^{-1}) + O_p((nh_1^{l_{1c}})^{-1})$ as in Theorem 1 by A4(1) and (3) and A6(1).

$$\begin{aligned} C_{62k} &= \frac{1}{n^2}\sum_t \sum_{i,t<i} \underbrace{[\phi(S_i,S_t) - \hat{E}\phi(S_i,S_t)]}_{\Phi(S_i,S_t)} + \frac{1}{n^2}\sum_t\sum_{i,t<i}\hat{E}\phi(S_i,S_t) \\ &= C_{621k} + C_{622k} \end{aligned}$$

where $\hat{E}\phi(S_i,S_t) = E(\phi(S_i,S_t)|S_t) + E(\phi(S_i,S_t)|S_i)$ since $E\phi(S_i,S_t) = 0$.
$C_{621k} = O(n^{-1}(h_2^{-\frac{(l_{1c}+l_{2c})}{2}} + h_1^{-\frac{l_{1c}}{2}}))$ as in Theorem 1 with A6(1).

$\sqrt{n}C_{622k} = \sqrt{n}\frac{1}{n^2}\sum_t\sum_{i,t<i}[E(\phi(S_i,S_t)|S_t) + E(\phi(S_i,S_t)|S_i)]$
$= \sqrt{n}\frac{1}{n^2}\sum_t\sum_{i,t\neq i} E(\phi(S_i,S_t)|S_t) = \frac{n-1}{n}\sum_{t=1}^n \frac{1}{\sqrt{n}}S_{tn} = \frac{n-1}{n}\sum_{t=1}^n \tilde{S}_{tn}$, where

$$\begin{aligned} S_{tn} &= \epsilon_t[\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}\sigma^2(Z_{1i})f_z(Z_i)} K_2(\frac{Z_{1t}^c - Z_{1i}^c}{h_2}, \frac{Z_{2t}^c - z_{2i}^c}{h_2}) I(Z_t^d = Z_i^d) f_z(Z_i) dZ_i \\ &\quad - \int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}\sigma^2(Z_{1i})f_1(Z_{1i})} K_1(\frac{Z_{1t}^c - Z_{1i}^c}{h_1}) I(Z_{1t}^d = Z_{1i}^d) f_z(Z_i) dZ_i]. \end{aligned}$$

Note that $\tilde{S}_{tn}$ forms am independent triangular array, $E\tilde{S}_{tn} = 0$, and

$$\begin{aligned} &\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}\sigma^2(Z_{1i})f_z(Z_i)} K_2(\frac{Z_{1t}^c - Z_{1i}^c}{h_2}, \frac{Z_{2t}^c - z_{2i}^c}{h_2}) I(Z_t^d = Z_i^d) f_z(Z_i) dZ_i \\ &\to \frac{g_k(Z_t) - g_{1,k}(Z_{1t})}{\sigma^2(Z_{1t})} \text{ uniformly } \forall Z_t \in G \text{ with A2(3) and (6) and A6(1),} \end{aligned}$$

$$\begin{aligned} &\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}\sigma^2(Z_{1i})f_1(Z_{1i})} K_1(\frac{Z_{1t}^c - Z_{1i}^c}{h_1}) I(Z_{1t}^d = Z_{1i}^d) f_z(Z_i) dZ_i \\ &\to E\frac{g_k(Z_{1t},Z_{2i}) - g_{1,k}(Z_{1t})}{\sigma^2(Z_{1t})} = 0 \text{ uniformly } \forall Z_{1t} \in G_1 \text{ with A2(3) and (6) and A6(1),} \end{aligned}$$

$\sum_{t=1}^n E\tilde{S}_{tn}^2 = V(S_{tn}) = E\frac{[g_k(Z_t) - g_{1,k}(Z_{1t})]^2}{\sigma^2(Z_{1t})} + o(1)$.
By Liapounov's Central Limit Theorem, provided $\lim_{n\to\infty}\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta} = 0$ for some $\delta > 0$, we have $\sum_t \tilde{S}_{tn} \xrightarrow{d} N(0, E\frac{[g_k(Z_t)-g_{1,k}(Z_{1t})]^2}{\sigma^2(Z_{1t})})$.
$\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta} = (\sum_t E\tilde{S}_{tn}^2)^{-1-\frac{\delta}{2}} n^{-\frac{\delta}{2}} E|S_{tn}|^{2+\delta}$,
$E|S_{tn}|^{2+\delta}$
$\leq c\{E[E(|\epsilon_t|^{2+\delta}|Z_t)|\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}\sigma^2(Z_{1i})f_z(Z_i)} K_2(\frac{Z_{1t}^c - Z_{1i}^c}{h_2}, \frac{Z_{2t}^c - z_{2i}^c}{h_2}) I(Z_t^d = Z_i^d) f_z(Z_i) dZ_i|^{2+\delta}]$
$+E[E(|\epsilon_t|^{2+\delta}|Z_t)|\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}\sigma^2(Z_{1i})f_1(Z_{1i})} K_1(\frac{Z_{1t}^c - Z_{1i}^c}{h_1}) I(Z_{1t}^d = Z_{1i}^d) f_z(Z_i) dZ_i|^{2+\delta}]\}$
Since $E(|\epsilon_t|^{2+\delta}|Z_t) < \infty$ by A4(3), and
$E|\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_2^{l_{1c}+l_{2c}}\sigma^2(Z_{1i})f_z(Z_i)} K_2(\frac{Z_{1t}^c - Z_{1i}^c}{h_2}, \frac{Z_{2t}^c - z_{2i}^c}{h_2}) I(Z_t^d = Z_i^d) f_z(Z_i) dZ_i|^{2+\delta}$
$\quad \to E|\frac{g_k(Z_t)-g_{1,k}(Z_{1t})}{\sigma^2(Z_{1t})}|^{2+\delta} \leq 2EX_{t,k}^{2+\delta}$, and similarly,
$E|\int \frac{g_k(Z_i)-g_{1,k}(Z_{1i})}{h_1^{l_{1c}}\sigma^2(Z_{1i})f_1(Z_{1i})} K_1(\frac{Z_{1t}^c - Z_{1i}^c}{h_1}) I(Z_{1t}^d = Z_{1i}^d) f_z(Z_i) dZ_i|^{2+\delta}$
$\quad \to E|E_{Z_{2i}|Z_{1t}}(\frac{g_k(Z_{1t},Z_{2i})-g_{1,k}(Z_{1t})}{\sigma^2(Z_{1t})})|^{2+\delta} = 0$,
So in all we have $E|S_{tn}|^{2+\delta} < \infty$ and $\lim_{n\to\infty}\sum_{t=1}^n E|\frac{\tilde{S}_{tn}}{(\sum_t E\tilde{S}_{tn}^2)^{\frac{1}{2}}}|^{2+\delta} = 0$.
Finally with Cramer-Rao device, for $C_6 = [C_{61}, C_{62}, \cdots, C_{6K}]'$, we obtain

$$\sqrt{n}C_6 \xrightarrow{d} N(0, E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t).$$

So combine results in (a) and (b), we obtain the claim in (1).

(2) (a) We first note since $\tilde{\beta} - \beta = O_p(n^{-\frac{1}{2}})$,

$$\begin{aligned} & \tilde{\epsilon}_t = m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}) + \epsilon_t - \hat{E}(\epsilon|Z_{1t}) + (X_t - \hat{E}(X|Z_{1t}))(\beta - \tilde{\beta}) \\ = \; & m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}) + \epsilon_t - \hat{E}(\epsilon|Z_{1t}) + (X_t - \hat{E}(X|Z_{1t}))O_p(n^{-\frac{1}{2}}) \end{aligned}$$

Since $X_t - \hat{E}(X|Z_{1t}) = X_t - E(X|Z_{1t}) + E(X|Z_{1t}) - \hat{E}(X|Z_{1t}) = X_t - E(X|Z_{1t}) + o_p(1)$ uniformly in $Z_{1t} \in G_1$, so consider the kth element in $X - E(X|Z_{1t})$, which is $X_k - E(X_k|Z_{1t})$,

$$\begin{aligned} & \hat{E}(X_k - E(X_k|Z_1)|Z_{1t}) \\ = \; & [\tfrac{f_1(Z_{1t}) - \hat{f}_1(Z_{1t})}{\hat{f}_1(Z_{1t})f_1(Z_{1t})} + \tfrac{1}{f_1(Z_{1t})}]\tfrac{1}{nh_1^{l_{1c}}}\textstyle\sum_{i=1}^n K_1(\tfrac{Z_{1i}^c - Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)(X_{i,k} - g_{1,k}(Z_{1i})) \\ = \; & O_p((\tfrac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O(h_1^{s+1}) \text{ as in Theorem 1, (1)(a) and (c).} \end{aligned}$$

Similarly $\hat{E}((X_k - E(X_k|Z_1))^2|Z_{1t})$
$= [o_p(1) + \frac{1}{f_1(Z_{1t})}]\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)(X_{i,k}^2 - 2x_{i,k}g_{1,k}(Z_{1i}) + g_{1,k}^2(Z_{1i}))$,
with assumption A6(3) we apply Lemma 1 and notice
$E\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - Z_{10}^c}{h_1})I(Z_{1i}^d = Z_{10}^d)(X_{i,k}^2 - 2x_{i,k}g_{1,k}(Z_{1i}) + g_{1,k}^2(Z_{1i}))$
$\rightarrow f_1(Z_{10})[E(X_k^2|Z_{10}) - g_{1,k}^2(Z_{10})] < \infty$ uniformly $\forall Z_{10} \in G_1$.
So we have $\hat{E}(X_k - E(X_k|Z_1)^2|Z_{1t}) = O_p(1)$ uniformly in Z_{1t}. We obtain

$$\begin{aligned} (b)\, \tilde{\epsilon}_t^2 \; & = (m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}))^2 + (\epsilon_t - \hat{E}(\epsilon|Z_{1t}))^2 + o_p(n^{-\frac{1}{2}}) \\ & + 2(m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}))(\epsilon_t - \hat{E}(\epsilon|Z_{1t})) + 2(m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}))O_p(n^{-\frac{1}{2}}) \\ & + 2(\epsilon_t - \hat{E}(\epsilon|Z_{1t}))O_p(n^{-\frac{1}{2}}) \end{aligned}$$

From result 2(b) and (f) in Theorem 1 proof, we have $2(m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}))O_p(n^{-\frac{1}{2}}) = o_p(n^{-\frac{1}{2}})$, $(m(Z_{1t}) - \hat{E}(m(z_1)|Z_{1t}))^2 = o_p(n^{-\frac{1}{2}})$, $2(\epsilon_t - \hat{E}(\epsilon|Z_{1t}))O_p(n^{-\frac{1}{2}}) = o_p(n^{-\frac{1}{2}})$, and $[\{\hat{E}(\epsilon|Z_{1i})]^2 = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-1}) = o_p(n^{-\frac{1}{2}})$ uniformly in Z_{1t}. So

$$\begin{aligned} (c) \quad & \hat{\sigma}^2(Z_{1t}) = \hat{E}(\tilde{\epsilon}^2|Z_{1t}) \\ = \quad & [o_p(1) + \tfrac{1}{f_1(Z_{1t})}]\tfrac{1}{nh_1^{l_{1c}}}\textstyle\sum_{i=1}^n K_1(\tfrac{Z_{1i}^c - Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)\{\epsilon_i^2 - 2\epsilon_i\hat{E}(\epsilon|Z_{1i}) + [\{\hat{E}(\epsilon|Z_{1i})]^2 \\ + \quad & 2(m(Z_{1i}) - \hat{E}(m(z_1)|Z_{1i}))(\epsilon_i - \hat{E}(\epsilon|Z_{1i}))\} + o_p(n^{-\frac{1}{2}}) \\ = \quad & [o_p(1) + \tfrac{1}{f_1(Z_{1t})}]\underbrace{\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K_1(\frac{Z_{1i}^c - Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)\epsilon_i^2}_{I} \\ + \quad & o_p(n^{-\frac{1}{2}}) + O_p((\tfrac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1^{s+1}) \end{aligned}$$

since $\frac{1}{nh_1^{l_{1c}}}\sum_{i=1}^n K(\frac{Z_{1i}^c - Z_{1t}^c}{h_1})I(Z_{1i}^d = Z_{1t}^d)|\epsilon_i| = O_p(1)$ with A6(1) and (2).

(d) With A6(3) and A4(4), we apply Lemma 1 to obtain $\sup_{Z_{1t} \in G_1}|I - EI| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})$. With a change of variable and using A6(1) and A2(1),

$$
\begin{aligned}
& EI = \int K_1(\Psi)\sigma^2(Z_{1t}^c + h_1\psi, Z_{1t}^d) f_1(Z_{1t}^c + h_1\psi, Z_{1t}^d) d\psi \\
= \;& \int K_1(\Psi)[\sigma^2(Z_{1t}) + \sum_{|j|=1}^s \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t})\tfrac{h_1^{|j|}\Psi^j}{j!} \\
+ \;& \sum_{|j|=s}(\tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}))\tfrac{h_1^{|j|}\Psi^j}{j!}][f_1(Z_{1t}) + \sum_{|l|=1}^s \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t})\tfrac{h_1^{|l|}\Psi^l}{l!} \\
+ \;& \sum_{|l|=s}(\tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}))\tfrac{h_1^{|l|}\Psi^l}{l!}]d\Psi \\
= \;& \sigma^2(Z_{1t})f_1(Z_{1t}) + \sigma^2(Z_{1t})\sum_{|j|=1}^s \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t})\tfrac{h_1^{|j|}}{j!}\int K(_1\Psi)\Psi^j d\Psi \\
+ \;& \sigma^2(Z_{1t})\sum_{|j|=s}\int K_1(\Psi)(\tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}))\tfrac{h_1^{|l|}\Psi^l}{l!}d\Psi \\
+ \;& \sum_{|j|=1}^s \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t})\tfrac{h_1^{|j|}}{j!}f_1(Z_{1t})\int K_1(\Psi)\Psi^j d\Psi \\
+ \;& \sum_{|j|=1}^s \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t})\tfrac{h_1^{|j|}}{j!}(\sum_{|l|=1}^s \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t})\tfrac{h_1^{|l|}}{l!}\int K_1(\Psi)\Psi^{j+l}d\Psi) \\
+ \;& \sum_{|j|=1}^s \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t})\tfrac{h_1^{|j|}}{j!}(\sum_{|l|=s}\tfrac{h_1^{|l|}}{l!}\int K_1(\Psi)(\tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}))\Psi^{j+l}d\Psi) \\
+ \;& f_1(Z_{1t})\sum_{|j|=s}\tfrac{h_1^{|j|}}{j!}\int K_1(\Psi)(\tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}))\Psi^j d\Psi \\
+ \;& \sum_{|j|=s}\sum_{|l|=1}^s \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t})\tfrac{h_1^{|l|}}{l!}\tfrac{h_1^{|j|}}{j!}\int K_1(\Psi)(\tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}))\Psi^{j+l}d\Psi \\
+ \;& \sum_{|j|=s}\sum_{|l|=s}\tfrac{h_1^{|l|}}{l!}\tfrac{h_1^{|j|}}{j!}\int K_1(\Psi)(\tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^l}{\partial (Z_{1t}^c)^l} f_1(Z_{1t})) \\
\times \;& (\tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}^c*, Z_{1t}^d) - \tfrac{\partial^j}{\partial (Z_{1t}^c)^j}\sigma^2(Z_{1t}))\Psi^{j+l}d\Psi \\
= \;& \sigma^2(Z_{1t})f_1(Z_{1t}) + O(h_1^{s+1}),\ \text{with the additional assumption A6(1).}
\end{aligned}
$$

The claim in (2) follows from (a)-(d).

$$
\begin{aligned}
(3) \quad & \sqrt{n}(\tilde{\beta}^I - \tilde{\beta}^H) \\
= \quad & \sqrt{n}\{[(\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1} - (\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1}]\hat{W}'\Omega^{-1}(\vec{Z}_1)(\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1)) \\
+ \quad & (\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1}\hat{W}'[\Omega^{-1}(\vec{Z}_1) - \hat{\Omega}^{-1}(\vec{Z}_1)](\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1))\}
\end{aligned}
$$

So we show

$$
\begin{aligned}
(a) \quad & (\tfrac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1} - (\tfrac{1}{n}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1} = o_p(1). \\
(b) \quad & \sqrt{n}\tfrac{1}{n}\hat{W}'[\Omega^{-1}(\vec{Z}_1) - \hat{\Omega}^{-1}(\vec{Z}_1)](\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1)) = o_p(1).
\end{aligned}
$$

Since in (1) we have $\sqrt{n}\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)(\hat{E}(Y|\vec{Z}) - \hat{E}(Y|\vec{Z}_1)) = O_p(1)$, and with (a) $(\frac{1}{n}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1} \xrightarrow{p} (E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1}$, $E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t$ is positive definite, the claim of (3) follows from (a) and (b).

We first note $\sup_{Z_{1t}\in G_1}|\frac{1}{\sigma^2(Z_{1t})} - \frac{1}{\hat{\sigma}^2(Z_{1t})}|$
$\leq [\inf_{Z_{1t}\in G_1}\sigma^2(Z_{1t})\inf_{Z_{1t}\in G_1}\hat{\sigma}^2(Z_{1t})]^{-1}\sup_{Z_{1t}\in G_1}|\hat{\sigma}^2(Z_{1t}) - \sigma^2(Z_{1t})|$.
With result (1) and A6(1), for large n, $\inf_{Z_{1t}\in G_1}\hat{\sigma}^2(Z_{1t}) > 0$, so

$$\sup_{Z_{1t}\in G_1}|\frac{1}{\sigma^2(Z_{1t})} - \frac{1}{\hat{\sigma}^2(Z_{1t})}| = O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}}) + O_p(h_1^{s+1}) + o_p(n^{-\frac{1}{2}}).$$

(a) Since $\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W} \xrightarrow{p} E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t$, which is positive definite, so by Slutsky's Theorem, $(\frac{1}{n}\hat{W}'\Omega^{-1}(\vec{Z}_1)\hat{W})^{-1} \xrightarrow{p} (E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1}$.
If (a') $\frac{1}{n}\hat{W}'(\Omega^{-1}(\vec{Z}_1) - \hat{\Omega}^{-1}(\vec{Z}_1))\hat{W} = o_p(1)$, then $\frac{1}{n}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W} \xrightarrow{p} E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t$ as well, and $(\frac{1}{n}\hat{W}'\hat{\Omega}^{-1}(\vec{Z}_1)\hat{W})^{-1} \xrightarrow{p} (E\frac{1}{\sigma^2(Z_{1t})}W_t'W_t)^{-1}$ so we have the claim in (a). So we only need to show (a').
The $(i,j)th$ element in $\frac{1}{n}\hat{W}'(\Omega^{-1}(\vec{Z}_1) - \hat{\Omega}^{-1}(\vec{Z}_1))\hat{W}$ is

$$
\begin{aligned}
& \tfrac{1}{n}\sum_{t=1}^{n}\hat{W}_{t,i}\left(\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right)\hat{W}_{t,j} \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][\hat{g}_i(Z_t)-g_i(Z_t)][\hat{g}_j(Z_t)-g_j(Z_t)] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][\hat{g}_i(Z_t)-g_i(Z_t)][\hat{g}_{1,j}(Z_{1t})-g_{1,j}(Z_{1t})] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][\hat{g}_i(Z_t)-g_i(Z_t)][\hat{g}_j(Z_t)-g_{1,j}(Z_{1t})] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_{1,i}(Z_{1t})-\hat{g}_{1,i}(Z_{1t})][\hat{g}_j(Z_t)-g_j(Z_t)] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_{1,i}(Z_{1t})-\hat{g}_{1,i}(Z_{1t})][\hat{g}_{1,j}(Z_{1t})-g_{1,j}(Z_{1t})] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_{1,i}(Z_{1t})-\hat{g}_{1,i}(Z_{1t})][\hat{g}_j(Z_t)-g_{1,j}(Z_{1t})] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_i(Z_t)-g_{1,i}(Z_{1t})][\hat{g}_j(Z_t)-g_j(Z_t)] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_i(Z_t)-g_{1,i}(Z_{1t})][\hat{g}_{1,j}(Z_{1t})-g_{1,j}(Z_{1t})] \\
= & \tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_i(Z_t)-g_{1,i}(Z_{1t})][\hat{g}_j(Z_t)-g_{1,j}(Z_{1t})] \\
= & A_1+\cdots+A_9
\end{aligned}
$$

Since $\sup_{Z_{1t}\in G_1}|\frac{1}{\sigma^2(Z_{1t})}-\frac{1}{\hat{\sigma}^2(Z_{1t})}|=o_p(1)$, we follow Theorem 1 (1) to have $A_i=o_p(1)$ for $i=1,\cdots,9$. So we have the claim in (a') and (a).

(b) The kth element in $\frac{1}{n}\hat{W}'[\Omega^{-1}(\vec{Z}_1)-\hat{\Omega}^{-1}(\vec{Z}_1)](\hat{E}(Y|\vec{Z})-\hat{E}(Y|\vec{Z}_1))$ is

$$
\begin{aligned}
C_k & =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][\hat{g}_k(Z_t)-g_k(Z_t)][\hat{E}(m(z_1)|Z_t)-\hat{E}(m(z_1)|Z_{1t})] \\
& =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_{1,k}(Z_{1t})-\hat{g}_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t)-\hat{E}(m(z_1)|Z_{1t})] \\
& =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_k(Z_t)-g_{1,k}(Z_{1t})][\hat{E}(m(z_1)|Z_t)-\hat{E}(m(z_1)|Z_{1t})] \\
& =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][\hat{g}_k(Z_t)-g_k(Z_t)][\hat{E}(\epsilon|Z_t)-\hat{E}(\epsilon|Z_{1t})] \\
& =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_{1,k}(Z_{1t})-\hat{g}_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t)-\hat{E}(\epsilon|Z_{1t})] \\
& =\tfrac{1}{n}\sum_t\left[\tfrac{1}{\sigma^2(Z_{1t})}-\tfrac{1}{\hat{\sigma}^2(Z_{1t})}\right][g_k(Z_t)-g_{1,k}(Z_{1t})][\hat{E}(\epsilon|Z_t)-\hat{E}(\epsilon|Z_{1t})]
\end{aligned}
$$

With $\sup_{Z_{1t}\in G_1}|\frac{1}{\sigma^2(Z_{1t})}-\frac{1}{\hat{\sigma}^2(Z_{1t})}|=O_p((\frac{nh_1^{l_{1c}}}{ln(n)})^{-\frac{1}{2}})+O_p(h_1^{s+1})+o_p(n^{-\frac{1}{2}})$, and Theorem 1 proof (1)(d), (f), (2)(b),(f), we easily have $\sqrt{n}C_k=o_p(1)$, thus we obtain the claim in (b).

References

[1] Ai, C. and X. Chen, 2003, Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica, 71, 1795-1843.

[2] Angrist, J. D., and A. B. Krueger, 1991, Does compulsory school attendance affect schooling and earnings. Quarterly Journal of Economics, 106, 979-1014.

[3] Angrist, J. D., and A. B. Krueger, 1995, Split-sample instrumental variables estimates of the return to schooling. Journal of Business & Economic Statistics, 13, 225-235.

[4] Baltagi, B. H. and Q. Li, 2002, On instrumental variable estimation of semiparametric dynamic panel data models. Economics Letters, 76, 1-9.

[5] Basmann, R. L., 1957, A generalized classical method of linear estimation of coefficients in a structural equation. Econometrica, 25, 77-83.

[6] Bickel, P., 1982, On adaptive estimation. Annals of Statistics, 10, 647-671.

[7] Blundell, R., and J. L. Powell, 2003, Endogeneity in semiparametric and nonparametric regression models, in Advances in economics and econometrics: theory and applications. Vol. 2, ed. by M. Dewatripont, L. P. Hansen, S. J. Turnovsky. Cambridge, U.K., Cambridge university press, 312-357.

[8] Blundell, R., X. Chen, and D. Kristensen, 2007, Semi-nonparametric IV estimation of shape-invariant Engle curves. Econometrica, 75, 1613-1669.

[9] Camlong-Viot, C., J. M. Rodríguez-Póo, and P. Vieu, 2006, Nonparametric and semiparametric estimation of additive models with both discrete and continuous variables under dependence. The art of semiparametrics, 155-178, Physica-Verlag HD.

[10] Card, D., 1995, Using geographic variation in college proximity to estimate the return to schooling, in Aspects of labour market behavior: essays in honour of John Vanderkamp. Ed. L. N. Christophides, E. K. Grant and R. Swidinsky, 201-222. Toronto: University of Toronto Press.

[11] Card, D., 2001, Estimating the return to schooling: progress on some persistent econometric problems. Econometrica, 69, 1127-1160,

[12] Chen, X. and D. Pouzo, 2012, Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. Econometrica, 80, 277-321.

[13] Chamberlain, G., 1992, Efficiency bounds for semiparametric regression. Econometrica, 60, 567-596.

[14] Darolles, S., Y. Fan, J.-P. Florens, and E. Renault, 2011, Nonparametric Instrumental Regression. Econometrica, 79, 1541-1566.

[15] Davidson, R., and J. G. MacKinnon, 2004, Econometric theory and methods. Oxford university press, New York.

[16] Deschenes, O., 2007, Estimating the effects of family background on the return to schooling. Journal of Business & Economic Statistics, 25, 265-277.

[17] Delgado, M.A., and J. Mora, 1995, Nonparametric and semiparametric estimation with discrete regressors. Econometrica, 63, 1477-1484.

[18] Fan, J., W. Härdle and E. Mammen, 1998, Direct estimation of low dimensional components in additive models. The Annals of Statistics, 26, 943-971.

[19] Florens, J.-P.,J. Johanns, and S. Van Bellegem, 2011, Identification and estimation by penalization in nonparametric instrumental regression. Econometric Theory, 27, 472-496.

[20] Griliches, Z., 1977, Estimating the returns to schooling: some econometric problems. Econometrica, 45, 1-22.

[21] Härdle, W., H. Liang, and J. Gao, 2000, Partially linear models. Physica-Verlag.

[22] Imbens, G. W. and W. K. Newey, 2009, Identification and estimation of triangular simultaneous equations models without additivity. Econometrica, 77, 1481-1512.

[23] Kling, J. R., 2001, Interpreting instrumental variables estimates of the returns to schooling. Journal of Business & Economic Statistics, 19, 358-364

[24] Lewbel, A., 2012, Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models. Journal of Business & Economic Statistics, 30, 67-80.

[25] Li, Q. and J. Racine, 2007, Nonparametric econometrics: theory and practice. Princeton University Press, Princeton, NJ.

[26] Li, Q., and T. Stengos, 1996, Semiparametric estimation of partially linear panel data models. Journal of Econometrics, 71, 389-397.

[27] Manski, C. F., 1984, Adaptive estimation of nonlinear regression models. Econometric Reviews, 3, 145-194.

[28] Martins-Filho, C. and F. Yao, 2007, Nonparametric frontier estimation via local linear regression. Journal of Econometrics, 141, 283-319.

[29] Matzkin, R.L.,1994, Restrictions of economic theory in nonparametric methods. In Handbook of Econometrics: IV, edited by R. F. Engle, and D. L. McFadden, Elsevier Science, North Holland.

[30] Matzkin, R.L., 2008, Identification in nonparametric simultaneous equations. Econometrica, 76, 945-978.

[31] Newey, W. K., 1990, Efficient instrumental variable estimation of nonlinear models. Econometrica, 58, 809-837.

[32] Newey, W. K., 1993, Efficient estimation of models with conditional moment restrictions. In Handbook of Statistics, v 11, 419-454, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., Amsterdam: North-Holland.

[33] Otsu, T., 2011. Empirical likelihood estimation of conditional moment restriction models with unknown functions. Econometric Theory 27, 8-46.

[34] Powell, J. L., J. H. Stock and T. M. Stoker, 1989, Semiparametric estimation of the index coefficients. Econometrica, 57, 1403-1430.

[35] Racine, J. and Q. Li, 2004, Nonparametric estimation of regression functions with both categorical and continuous data. Journal of Econometrics, 119, 99-130.

[36] Robinson, P., 1988, Root-n-consistent semiparametric regression. Econometrica, 56, 931-954.

[37] Speckman, P., 1988, Kernel smoothing in partial linear models. Journal of the Royal Statistical Society, Series B, 50, 413-436.

[38] Su, L., I. Murtazashvili, and A. Ullah, 2011, Local linear GMM estimation of functional coefficient IV models with an application to estimating the rate of return to schooling, manuscript, Singapore Management University.

[39] Ruppert, D., S. J. Sheather, M. P. Wand, 1995, An effective bandwidth selector for local least squares regression. Journal of the American Statistical Association, 90, 1257-1270.

[40] Theil, H., 1953, Repeated least squares applied to complete equation systems. The Hague: central planning bureau.

[41] White, H., 1980, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, 48, 817-838.

[42] White, H., 1982, Instrumental variables regression with independent observations. Econometrica, 50, 483-499.

[43] Yao, F. and J. Zhang, 2012, Efficient semiparametric instrumental variable estimation, working paper, Economics Department, West Virginia University. http://community.wvu.edu/~fy006/hp/yao-zhang2012.pdf