NBER WORKING PAPER SERIES

REVISITING INSTRUMENTAL VARIABLES AND THE CLASSIC CONTROL FUNCTION APPROACH, WITH IMPLICATIONS FOR PARAMETRIC AND NON-PARAMETRIC REGRESSIONS

Kyoo il Kim
Amil Petrin

Working Paper 16679
http://www.nber.org/papers/w16679

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2011

Revisiting Instrumental Variables and the Classic Control Function Approach, with Implications for Parametric and Non-Parametric Regressions
Kyoo il Kim and Amil Petrin
NBER Working Paper No. 16679
January 2011
JEL No. C14

ABSTRACT

We show that the well-known numerical equivalence between two-stage least squares (2SLS) and the classic control function (CF) estimator raises an interesting and unrecognized puzzle. The classic CF approach maintains that the regression error is mean independent of the instruments conditional on the CF control, which is not required by 2SLS, and could easily be violated. We show that the classic CF estimator can be modified to allow the mean of the error to depend in a general way on the instruments and control by adding the unconditional moment restrictions maintained by 2SLS. In this case 2SLS and our generalized CF estimator are no longer numerically equivalent, although asymptotically both converge to the true value. We then show that our generalized CF estimator is consistent in parametric or non-parametric settings with endogenous regressors and additive errors. For example, our estimator is consistent when the conditional mean of the error depends on the instruments while the nonparametric estimator of Newey, Powell, and Vella (1999) based on the classic CF restriction is not. Our new approach is also not subject to the ill-posed inverse problem that affects the non-parametric estimator of Newey and Powell (2003). Our estimator is easy to implement in standard programming packages - it is a multi-step least squares estimator - and our monte carlos show that our new estimator performs well while the classical CF estimator and the non-parametric analog of Newey, Powell, and Vella (1999) can be biased in non-linear settings when the conditional mean of the error depends on the instruments.

Kyoo il Kim
Department of Economics
University of Minnesota
4-129 Hanson Hall
1925 4th Street South
Minneapolis, MN 55455
kyookim@umn.edu

Amil Petrin
Department of Economics
University of Minnesota
4-101 Hanson Hall
Minneapolis, MN 55455
and NBER
petrin@umn.edu

1 Introduction

The problem of endogenous regressors in simultaneous equations models has a long history in econometrics and empirical studies. In models with additively separable errors researchers have used both the instrumental variable (IV) approach and the classical control function (CF) approach to correct for the bias induced by the correlation between the error and the regressor(s).[1] While both approaches yield consistent estimates if their assumptions hold, they differ in exactly the assumptions they maintain. In addition, both the IV and CF's maintained assumptions vary depending upon whether the model is linear, nonlinear,[2] or non-parametric in the regressors (see the review in Blundell and Powell (2003)).

We start this paper by developing new results on the relationship between the IV and CF estimators. We start with the base case, where the principal equation of interest is linear in regressors. We call this equation the structural equation. For identification both estimators require sufficient exclusion restrictions (or order conditions) and their associated moment conditions. The classic CF estimator further imposes that the first moment of the error in the structural equation cannot depend in any way on exogenous variables, conditional on the standard CF control (i.e. the mean projection residual obtained from regressing the endogenous variable on the instruments). Since it is well known that the two stage least squares estimator (2SLS) and the classic CF estimator are numerically identical, the question arises as to whether this latter assumption is necessary for consistency of the CF estimator. We show that it is not - as it could not be given the equivalence result - although the estimated control function is no longer consistent for the expected value of the error conditional on the control and exogenous variables.[3] Furthermore, the classic CF estimator can be generalized to allow the expected value of the error to depend on both the classic CF control and other exogenous variables by adding the moment restrictions used by 2SLS for identification. In this case, 2SLS and the generalized CF estimator are no longer numerically equivalent, although they are both consistent.

We then turn to the non-linear and the non-parametric setting. We show how to use our insights from the linear case to develop a new and simple multi-step least squares estimator for non-linear and non-parametric models. This new estimator combines the strengths of

[1] For the classic control function approach see, for example, Telser (1964), Hausman (1978), or Heckman (1978).

[2] We use "nonlinear model" to refer to a regression model that is nonlinear in regressors but linear in parameters.

[3] In the classic CF case this is typically viewed as a nuisance parameter, although sometimes it is used to test for endogeneity. The IV estimator does not estimate the expected value of the error.

both the non-parametric 2SLS (NP2SLS) estimator of Newey and Powell (2003) and the non-parametric CF (NPVCF) estimator of Newey, Powell, and Vella (1999) while avoiding a weakness of each of these approaches. Specifically, while we utilize the same assumptions for identification from Newey and Powell (2003) (a conditional mean restriction and a sufficient order condition, (e.g.) a completeness condition), by adding a "generalized CF" moment condition we get an estimator that does not suffer from the ill-posed inverse problem faced by the NP2SLS estimator and the associated complications that arise.[4] Our new estimator also does not require the NPVCF condition that the expected value of the error in the structural equation must not depend on exogenous variables. After developing convergence rates for our estimators and consistent estimators for the standard errors of our multi-step estimator, we then provide several monte carlos that illustrate the ease of implementing our estimator. The monte carlos also show that our estimator remains consistent when the NPVCF condition previously noted does not hold, while the NPVCF becomes inconsistent, as theory suggests it would.

The paper proceeds as follows. In the next section we consider the linear additive model. Section 3 uses the results from Section 2 to formulate our new estimator for the non-linear or non-parametric setting. Section 4 discusses identification and Section 5 develops the details of our estimator. Section 6 addresses convergence rates and Section 7 provides conditions under which asymptotic normality holds for several structural objects often of interest Section 8 provides monte carlos and Section 9 concludes.

2 The Linear Setting with Additive Errors

Our first set of results will relate the IV and CF estimators in the base case, the linear setting with separable errors. We start by reviewing the 2SLS estimator and the classic control function estimator for the linear simultaneous equations model in mean-deviated form,

$$y_i \quad = \quad x_i\beta_0 + \varepsilon_i, \tag{1}$$

with y_i the dependent variable and x_i a scalar explanatory variable that is potentially correlated with ε_i, so $E[x_i\varepsilon_i] \neq 0$ (and $E[y_i] = 0$ and $E[x_i] = 0$). We let z_i denote an instrument vector satisfying

$$E[z_i\,\varepsilon_i] = 0, \quad E[z_i\,x_i] \neq 0, \tag{2}$$

[4] These complications include the question of existence of a unique solution and the difficulties associated with computation and estimation.

(and $E[z_i] = 0$). $\hat{\beta}_{2SLS}$, the 2SLS estimator for β_0, is calculated by first projecting x_i on the instruments z_i to recover the predicted values of x_i given z_i, which we denote $\hat{x}_i$. y_i is then regressed on $\hat{x}_i$ to get $\hat{\beta}_{2SLS}$. $\hat{\beta}_{CF}$, the control function estimator, is also calculated in two steps, with the first step projecting x_i off of the instruments z_i to recover the mean projection residual $\hat{v}_i = x_i - \hat{x}_i$. The second step regresses y_i on x_i and $\hat{v}_i$, with the coefficient on x_i from this regression the estimator for $\hat{\beta}_{CF}$.

It is well known that these two approaches yield numerically identical estimates. We provide a simple proof using projection that shows why they are equivalent. Let $Y = (y_1, \ldots, y_n)'$, $X = (x_1, \ldots, x_n)'$, $Z = (z_1, \ldots, z_n)'$, $\hat{X} = (\hat{x}_1, \ldots, \hat{x}_n)'$, and $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_n)'$.

Theorem 1 (2SLS-CF Numerical Equivalence)**.** *If*

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'Y \text{ and}$$

$$(\hat{\beta}_{CF}, \hat{\rho}_{CF}) = ((X, \hat{V})'(X, \hat{V}))^{-1}(X, \hat{V})'Y$$

are well-defined and exist, then $\hat{\beta}_{2SLS} = \hat{\beta}_{CF}$.

Proof. From projection theory the same numerical estimate obtains for the coefficient on x_i from either regressing Y on $(X, \hat{V})$ or regressing Y on the projection of X off of $\hat{V}$. Numerical equivalence follows because the projection of X off of $\hat{V}$ is equal to $\hat{X}$ because $(I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}')X = (I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}')(\hat{X} + \hat{V}) = \hat{X}$, as $\hat{V}'\hat{X} = 0$ by projection. □

While the proof is helpful for understanding the numerical equivalence of these two approaches, it masks the fact that the control function estimator places a different restriction on the data generating process relative to 2SLS. We recast the two estimators in terms of the conditional expectations they use to illustrate this difference.

In the first step of 2SLS an estimate of $E[x_i|z_i]$ is constructed using projection, and in the second stage the dependent variable y_i is regressed on the estimate of $E[x_i|z_i]$ to obtain $\hat{\beta}_{2SLS}$. Even though this new regressor measures x_i with error, the error is by construction uncorrelated with the new regressor $\hat{x}_i$ leading 2SLS to be consistent.

Defining $v_i = x_i - E[x_i|z_i]$, the spirit of the classic CF estimator is to regress y_i on $E[x_i|z_i, v_i] = E[x_i|z_i, x_i] = x_i$ and $E[\varepsilon_i|z_i, v_i]$ (or $E[\varepsilon_i|z_i, x_i]$), where conditioning the error on (z_i, v_i) controls for its correlation with x_i, leaving the remaining variation in x_i exogenous. In implementation the classic CF approach estimates the following equation:

$$y_i = x_i\beta + \rho v_i + \eta_i, \tag{3}$$

replacing v_i with $\hat{v}_i$, thus imposing

$$(\textbf{CF Restriction})\ E[\varepsilon_i|z_i,x_i] = E[\varepsilon_i|z_i,v_i] = E[\varepsilon_i|v_i], \tag{4}$$

and we refer to this as the classic CF restriction (relative to 2SLS).[5] The first equality requires that v_i be chosen such that, conditional on it and z_i, x_i is known, which is satisfied by the way v_i is defined. The second equality insists that the mean of ε_i does not depend on the instruments z_i conditional on the control v_i, a restriction that 2SLS does not impose. Our first main result of interest arises from this apparent "puzzle"; $\hat{\beta}_{2SLS} = \hat{\beta}_{CF}$, *but the classic CF approach maintains a different stochastic assumption on* ε_i *relative to 2SLS that could easily be violated.*

We resolve this puzzle in two steps. We start by considering an (unrestricted) general specification for the conditional expectation of the error

$$E[\varepsilon_i|z_i,v_i] \equiv h(z_i,v_i) = \tilde{\rho}v_i + \tilde{h}(z_i,v_i),$$

with the function characterizing $E[\varepsilon_i|z_i,v_i]$ having a leading term in v_i and a remaining term denoted by the function $\tilde{h}(z_i,v_i)$. We first show that the instruments are uncorrelated with $\tilde{h}(z_i,v_i)$ when $E[z_i\,\varepsilon_i] = 0$ and v_i is constructed as the classic control function variable.

Theorem 2. *If* $E[z_i\,\varepsilon_i] = 0$ *and* v_i *is such that* $E[v_i|z_i] = 0$ *then* $E[z_i\,\tilde{h}(z_i,v_i)] = 0$.

Proof. Using the law of iterated expectations we have

$$\begin{aligned} 0 &= E[z_i\,\varepsilon_i] = E[z_i\,E[E[\varepsilon_i|z_i,v_i]|z_i]\,] = E[z_i(\tilde{\rho}E[v_i|z_i] + E[\tilde{h}(z_i,v_i)|z_i])] \qquad (5)\\ &= E[z_iE[\tilde{h}(z_i,v_i)|z_i]] = E[z_i\,\tilde{h}(z_i,v_i)]. \end{aligned}$$

□

We then consider the generalization of equation (3), with

$$Y = X\beta_0 + \tilde{\rho}\hat{V} + \tilde{H}(Z,\hat{V}) + \hat{\eta} \tag{6}$$

where $\tilde{H}(Z,\hat{V}) = (\tilde{h}(z_1,\hat{v}_1),\ldots,\tilde{h}(z_n,\hat{v}_n))'$ and $\hat{V} = X - Z\hat{\pi}$, the vector of controls, which are the ordinary least squares residuals from the regression of X on Z, and $\hat{\eta}$ is the residual after the estimated vector of controls are included. We can rewrite (6) as

$$Y = (Z\hat{\pi} + \hat{V})\beta_0 + \tilde{\rho}\hat{V} + \tilde{H}(Z,\hat{V}) + \hat{\eta}. \tag{7}$$

[5] From (3) it is also clear that in the implementation the classic CF approach further assumes $E[\varepsilon_i|v_i] = \rho v_i$.

Define $M_{\hat{V}} = I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}'$, the matrix that projects off of $\hat{V}$, and note that $M_{\hat{V}}\hat{V} = 0$ and $M_{\hat{V}}Z = Z$ (because $\hat{V}'Z = 0$). Then by partitioned regression theory, estimation of β_0 in (7) is numerically equivalent to the estimation of

$$\begin{aligned} Y &= M_{\hat{V}}(Z\hat{\pi} + \hat{V})\beta_0 + M_{\hat{V}}\tilde{H}(Z, \hat{V}) + M_{\hat{V}}\hat{\eta} \\ &= Z\hat{\pi}\beta + M_{\hat{V}}\tilde{H}(Z, \hat{V}) + M_{\hat{V}}\hat{\eta}. \end{aligned}$$

If $M_{\hat{V}}\tilde{H}(Z, \hat{V})$ is asymptotically uncorrelated with $Z\hat{\pi}$, i.e., if $Z'M_{\hat{V}}\tilde{H}(Z, \hat{V})/n$ converges to zero as the sample size goes to infinity, then the least squares estimator of β_0 in (6) is consistent whether or not we include $\tilde{H}(Z, \hat{V})$ in the regression as long as $\hat{V}$ is included as a regressor in (6).

Theorem 3 couples (2) with weak regularity conditions which are sufficient for $Z'M_{\hat{V}}\tilde{H}(Z, \hat{V})/n$ to converge to zero.

Theorem 3. *Assume (i) $E[\|z_i\| \cdot ||\tilde{h}(z_i, v_i)||] < \infty$, (ii) $\tilde{h}(z, v)$ is differentiable with respect to v, (iii) for $v_i(\pi) \equiv x_i - z_i'\pi$, assume $\sup_{\pi^* \in \Pi_0} E[\|z_i\|^2 \left\|\frac{\partial \tilde{h}(z_i, v_i(\pi^*))}{\partial v_i}\right\|] < \infty$ for Π_0 some neighborhood of π_0, (iv) assume $E[\|z_i\|^2 \left\|\frac{\partial \tilde{h}(z_i, v_i(\pi))}{\partial v_i}\right\|]$ is continuous at $\pi = \pi_0$, and (v) $\hat{\pi} \to_p \pi_0$. If (2) holds then $Z'M_{\hat{V}}\tilde{H}(Z, \hat{V})/n \to 0$ as $n \to \infty$.*

Proof. We can reexpress as

$$Z'M_{\hat{V}}\tilde{H}(Z, \hat{V})/n = Z'(I - \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}')\tilde{H}(Z, \hat{V})/n = Z'\tilde{H}(Z, \hat{V})/n = \sum_{i=1}^{n} z_i\tilde{h}(z_i, \hat{v}_i)/n$$

because $Z'\hat{V} = 0$. Write $\sum_{i=1}^n z_i\tilde{h}(z_i, \hat{v}_i)/n = \sum_{i=1}^n z_i\tilde{h}(z_i, v_i)/n + \sum_{i=1}^n z_i(\tilde{h}(z_i, \hat{v}_i) - \tilde{h}(z_i, v_i))/n$. We have $\sum_{i=1}^n z_i\tilde{h}(z_i, v_i)/n \to_p E[z_i\tilde{h}(z_i, v_i)]$ by the law of large numbers under (i). Obtain $||\sum_{i=1}^n z_i(\tilde{h}(z_i, \hat{v}_i) - \tilde{h}(z_i, v_i))/n|| \leq \|\hat{\pi}^* - \pi_0\| \sum_{i=1}^n \|z_i\|^2 ||\frac{\partial \tilde{h}(z_i, v_i(\hat{\pi}^*))}{\partial v_i}||/n$ by applying the mean-value expansion, where $\hat{\pi}^*$ lies between $\hat{\pi}$ and π_0 and $v_i(\pi) = x_i - z_i'\pi$. Then the term $\sum_{i=1}^n z_i(\tilde{h}(z_i, \hat{v}_i) - \tilde{h}(z_i, v_i))/n \to_p 0$ by the consistency of $\hat{\pi}$ and $\sum_{i=1}^n \|z_i\|^2 \left\|\frac{\partial \tilde{h}(z_i, v_i(\hat{\pi}^*))}{\partial v_i}\right\|/n \to_p E[\|z_i\|^2 \left\|\frac{\partial \tilde{h}(z_i, v_i(\pi_0))}{\partial v_i}\right\|] < \infty$ under (iii) and (iv). Therefore $\sum_{i=1}^n z_i\tilde{h}(z_i, \hat{v}_i)/n \to_p E[z_i\tilde{h}(z_i, v_i)] = 0$ by (2) and Theorem 2. □

Several results emerge from these theorems. The main one is that the CF estimator is consistent whether or not we include the term $\tilde{H}(Z, \hat{V})$ in the estimation as long as we include the control $\hat{V}$ in (6). Thus the classic CF estimator based only on including $\hat{V}$ is consistent under (2), the same stochastic assumption required for the consistency of the 2SLS estimator. If the classic CF estimator is modified to include the new regressors associated with $\tilde{H}(Z, \hat{V})$ - our generalized CF estimator - then 2SLS and this generalized CF estimator for β_0 are no longer numerically equivalent, although asymptotically they both converge to β_0.

The theorems also make it clear that the classic control function approach does *not* generally yield consistency for the expected value of the error conditional on the control and exogenous variables unless $\tilde{H}(Z, \hat{V})$ is also included in the regression equation. Although this is not typically the object of interest of either the classic CF estimator or the 2SLS estimator, an exception is when one tests for endogeneity based on the estimate of ρ in (3).[6]

A simple example is illustrative of these points. Consider the case where z_i is a scalar and

$$E[\varepsilon_i | z_i, v_i] = \rho_1 v_i + v_i z_i \rho_2, \tag{8}$$

but the researcher only includes x_i and $\hat{v}_i$ as regressors. Even though the researcher omits the *relevant* variable $\hat{v}_i z_i$, the ordinary least squares estimator $\hat{\beta}$ is consistent for β_0 because

$$\sum_{i=1}^{n} \hat{v}_i z_i^2 / n \rightarrow_p E[v_i z_i^2] = 0,$$

which follows from $\hat{v}_i \rightarrow_p v_i$ (because $\hat{\pi} \rightarrow_p \pi_0$) and $E[v_i | z_i] = 0$ and by LLN under standard regularity conditions ($E[\|v_i\| \cdot \|z_i\|^2] < \infty$ and $E[\|z_i\|^3] < \infty$). However, $\hat{\rho}_1 \hat{v}_i$ is not a consistent estimator of $E[\varepsilon_i | z_i, v_i]$. If one desired a consistent estimator of this conditional expectation, then $\hat{v}_i z_i$ would have to be included in the regression, and $\hat{\rho}_1 \hat{v}_i + \hat{\rho}_2 \hat{v}_i z_i$ would be consistent for $E[\varepsilon_i | z_i, v_i]$. In this case 2SLS and our generalized CF estimates for β_0 would not longer be numerically equivalent in a finite sample, although they would both converge asymptotically to β_0.

[6] See e.g.Smith and Blundell (1986). In this case the misspecification of this conditional expectation may reduce the power of the test or call into question the test's consistency.

3 The Non-Linear or Non-Parametric Setting with Additive Errors

We consider a nonparametric simultaneous equations model with additivity:

$$x_i = \Pi_0(z_i) + v_i \tag{9}$$

$$y_i = f_0(x_i, z_{1i}) + \varepsilon_i. \tag{10}$$

where the instruments z_i includes z_{1i} and$f(x_i, z_{1i})$ can be parametric as $f(x_i, z_{1i}) \equiv f(x_i, z_{1i}; \theta)$ or nonparametric. We strengthen the unconditional moment condition to the conditional moment condition (CMR)

$$(\mathbf{CMR}) \qquad E[\varepsilon_i | z_i] = 0,$$

as we must in the non-parametric setting for identification. In this setting equation (9) is not a "structural" equation but simply an orthogonal decomposition, so $E[v_i|z_i] = 0$ is not restrictive.

Two approaches to estimation exist for an equation of the form (10). The non-parametric 2SLS (NP2SLS) estimator of Newey and Powell (2003) solves the integral equation implied by the **CMR** condition

$$E[y|z] = E[f_0(x, z_1)|z] = \int f_0(x, z_1)\mu(dx|z)$$

where μ denotes the conditional c.d.f. of x given z. While evidently the natural approach if the CMR condition is maintained, it suffers from the well-known ill-posed inverse problem.

The second approach is the non-parametric control function estimator of Newey, Powell, and Vella (1999) (NPVCF), which does not use the CMR condition but instead uses the orthogonal decomposition from equation (9) and maintains the classic CF restriction from (3) (i.e. $E[\varepsilon_i|z_i, v_i] = E[\varepsilon_i|v_i]$). Identification of $f(x_i, z_{1i})$ for the NPVCF estimator is thus achieved by ruling out the possibility that the error has an additive functional relationship with (x_i, z_{1i}). As noted in Section 2, this assumption is not necessarily innocuous, and while unnecessary in the linear case, is necessary for identification of the NPVCF estimator.[7]

As we note previously (9) itself is not restrictive because it does not need to be the true genesis of the endogenous regressor x_i. However, the mean independence restriction $E[\varepsilon_i|z_i, v_i] = E[\varepsilon_i|v_i]$ is restrictive. To understand this point suppose $x_i = r_0(z_i, v_i^*)$ is the true genesis of x_i where v_i^* is possibly a vector. Then even when ε_i is independent with z_i given the *true* control v_i^*, it is possible that $E[\varepsilon_i|z_i, v_i] \neq E[\varepsilon_i|v_i]$ if we use the *pseudo*

[7] For example, let $v_i = \psi(z_i)\varepsilon_i$ ($\psi(z_i) \neq 0$). Then we have $\varepsilon_i = v_i/\psi(z_i)$ and $E[\varepsilon_i|z_i, v_i] = v_i/\psi(z_i) \neq E[\varepsilon_i|v_i]$ unless $\psi(z_i)$ is constant.

control v_i from (9), so the classic CF restriction essentially requires that (9) should be the true genesis of x_i, the true structural equation. On the other hand in our approach we can still use the *pseudo* control v_i regardless of the true genesis of x_i as we develop our estimator in the next section.

3.1 The Conditional Moment Restriction-Control Function (CMRCF) estimator

We now describe our new estimator, which combines the strengths of both the NP2SLS estimator and the NPVCF estimator while avoiding a key weakness of each of these approaches. We consider the following regression, which is based on our generalized version of the classic CF estimator from Section 2

$$y_i = f_0(x_i, z_{1i}) + h_0(z_i, v_i) + \eta_i, \quad with \ \ E[\eta_i|z_i, v_i] = 0. \tag{11}$$

Without further restrictions on $h_0(z_i, v_i)$, $f_0(x_i, z_{1i})$ is not identified because $h_0(z_i, v_i)$ can be a function of (x_i, z_{1i}). To achieve identification we add to this model the **CMR** condition, which implies that the function $h_0(z_i, v_i)$ must satisfy $E[h_0(z_i, v_i)|z_i] = 0$, because by the law of iterated expectations

$$0 = E[\varepsilon_i|z_i] = E[E[\varepsilon_i|z_i, v_i]|z_i] = E[h_0(z_i, v_i)|z_i]. \tag{12}$$

This restriction will suffice for identification of $f_0(x_i, z_{1i})$ as we show below. Our approach thus loosens the classic **CF** restriction in (4) by combining the generalized CF moment in (11) with the more commonly used **CMR** restriction, and we refer to our estimator as the **CMRCF** estimator.[8]

We discuss general identification in Section 4 and we formalize our estimation procedure as a sieve method in Section 5. Here we provide a simple example that is illustrative of how in general identification is obtained, that is, how we can identify $f_0(x_i, z_{1i})$ from an additive regression of y_i on (x_i, z_{1i}) and (z_i, v_i) where (z_i, v_i) enters the control function $h_0(z_i, v_i)$ satisfying the **CMR** condition. Conditional on (z_i, v_i), the expectation of y_i (from (10)) is equal to

$$E[y_i|z_i, v_i] = f_0(x_i, z_{1i}) + E[\varepsilon_i|z_i, v_i] \equiv f_0(x_i, z_{1i}) + h_0(z_i, v_i) \tag{13}$$

because x_i is known given z_i and v_i. For this example we assume

$$h_0(z_i, v_i) = a_1(\Pi_0(z_i) + v_i) + a_2 v_i^2 + a_3' z_i v_i + \varphi(z_i) = a_1 x_i + a_2 v_i^2 + a_3' z_i v_i + \varphi(z_i)$$

[8]Note that the classic CF restriction does not imply the CMR restriction and vice versa.

where $\varphi(z_i)$ denotes any arbitrary function of z_i. Then the **CMR** condition implies that

$$\begin{aligned} 0 &= E[h_0(z_i, v_i)|z_i] = a_1 E[x_i|z_i] + a_2 E[v_i^2|z_i] + a_3' E[z_i v_i|z_i] + E[\varphi(z_i)|z_i] \\ &= a_1 \Pi_0(z_i) + a_2 E[v_i^2|z_i] + \varphi(z_i) \end{aligned}$$

since $E[v_i|z_i] = 0$. It follows that

$$\begin{aligned} & h_0(z_i, v_i) = h_0(z_i, v_i) - E[h_0(z_i, v_i)|z_i] \\ & = a_1 v_i + a_2(v_i^2 - E[v_i^2|z_i]) + a_3' z_i v_i + (\varphi(z_i) - \varphi(z_i)) = a_1 v_i + a_2 \tilde{v}_{2i} + a_3 z_i v_i \end{aligned}$$

where $\tilde{v}_{2i} = v_i^2 - E[v_i^2|z_i]$. Thus the **CMR** condition puts shape restrictions on $h_0(z_i, v_i)$ so it is not a function of x_i and it does not contain functions of z_i only. Identification in this example is then equivalent to the non-existence of a linear functional relationship among $x_i, z_{1i}, v_i, \tilde{v}_{2i}$, and $z_i v_i$ (i.e. no perfect multicollinearity).

Estimation proceeds in three simple steps. In the first step we obtain the control $\hat{v}_i = x_i - \hat{E}[x_i|z_i]$ from the first stage nonparametric regression (e.g., series estimation in Newey (1997) or sieve estimation in Chen (2007)). In the second step we construct an approximation of $h(z_i, \hat{v}_i)$ using (e.g.) polynomial approximations while imposing the restriction $E[h(z_i, v_i)|z_i] = 0$. For example, we can take

$$h(z_i, \hat{v}_i) \approx \sum\nolimits_{l_1=1}^{L_1} a_{l_1,0}(\hat{v}_i^{l_1} - E[\hat{v}_i^{l_1}|z_i]) + \sum\nolimits_{l=2}^{L} \sum_{l_1 \geq 1, l_2 \geq 1 \text{ s.t. } l_1 + l_2 = l} a_{l_1,l_2} \varphi_{l_2}(z_i)(\hat{v}_i^{l_1} - E[\hat{v}_i^{l_1}|z_i])$$

where $\varphi_{l_2}(z_i)$ denotes functions of z_i, $L_1, L \to \infty$ and $L_1/n, L/n \to 0$ as $n \to \infty$, and we approximate $E[\hat{v}_i^{l_1}|z_i]$ using (possibly nonparametric) regressions. In the last step we estimate $f(x_i, z_{1i})$ by including $h(z_i, \hat{v}_i)$ in the regression, estimating $f(x_i, z_{1i})$ (or θ in the parametric function $f(x_i, z_{1i}; \theta)$) and $h(z_i, \hat{v}_i)$ simultaneously.

4 Identification

We ask whether $f_0(x_i, z_{1i})$ is identified by equation (11) with restrictions (12). Our approach to identification closely follows Newey, Powell, and Vella (1999) and Newey and Powell (2003). We consider pairs of functions $\bar{f}(x_i, z_{1i})$ and $\bar{h}(z_i, v_i)$ that satisfy the conditional expectation in (13) and (12). Because conditional expectations are unique with probability one, if there is such a pair $\bar{f}(x_i, z_{1i})$ and $\bar{h}(z_i, v_i)$, it must be that

$$\Pr(f_0(x_i, z_{1i}) + h_0(z_i, v_i) = \bar{f}(x_i, z_{1i}) + \bar{h}(z_i, v_i)) = 1. \tag{14}$$

Identification of $f_0(x_i, z_{1i})$ means we must have $f_0(x_i, z_{1i}) = \bar{f}(x_i, z_{1i})$ whenever (14) holds. Working with differences, we let $\delta(x_i, z_{1i}) = f_0(x_i, z_{1i}) - \bar{f}(x_i, z_{1i})$ and $\kappa(z_i, v_i) = h_0(z_i, v_i) -$

$\bar{h}(z_i, v_i)$, with $E[\kappa(z_i, v_i)|z_i] = 0$ by (12). Identification of $f_0(x_i, z_{1i})$ is then equivalent to

$$\Pr(\delta(x_i, z_{1i}) + \kappa(z_i, v_i) = 0) = 1 \text{ implying } \Pr(\delta(x_i, z_{1i}) = 0, \kappa(z_i, v_i) = 0) = 1.$$

Theorem 4 (Identification with CMR)**.** *If equations (11) and (12) are satisfied, then $f_0(x_i, z_{1i})$ is identified if for all $\delta(x_i, z_{1i})$ with finite expectation, $E[\delta(x_i, z_{1i})|z_i] = 0$ implies $\delta(x_i, z_{1i}) = 0$ a.s.*

Proof. Suppose it is not identified. Then we must find functions $\delta(x_i, z_{1i}) \neq 0$ and $\kappa(z_i, v_i) \neq 0$ with $E[\kappa(z_i, v_i)|z_i] = 0$ such that $\Pr(\delta(x_i, z_{1i}) + \kappa(z_i, v_i) = 0) = 1$. But this is not possible because $0 = E[\delta(x_i, z_{1i}) + \kappa(z_i, v_i)|z_i] = E[\delta(x_i, z_{1i})|z_i]$ and $E[\delta(x_i, z_{1i})|z_i] = 0$ implies $\delta(x_i, z_{1i}) = 0$ a.s., so $\Pr(\delta(x_i, z_{1i}) = 0, \kappa(z_i, v_i) = 0) = 1$. □

We consider several cases, with the regressors demeaned in each example. For the simple model $f_0(x_i, z_{1i}) = \beta_0 x_i$, we have the alternative function $\tilde{f}(x_i, z_{1i}) = \tilde{\beta} x_i \neq \beta_0 x_i$. We have $\delta(x_i, z_{1i}) = (\beta_0 - \tilde{\beta})x_i$, so $E[\delta(x_i, z_{1i})|z_i] = 0$ implies $\delta(x_i, z_{1i}) = 0$ (or $(\beta_0 = \tilde{\beta})$) as long as $E[x_i|z_i] \neq 0$. Identification is then equivalent to z_i being correlated x_i, the standard instrumental variable condition.

The general case is given by $f_0(x_i, z_{1i}) = \beta_0' x_i + \beta_{10}' z_{1i}$. An alternative function is $\tilde{f}(x_i, z_{1i}) = \tilde{\beta}' x_i + \tilde{\beta}_1' z_{1i} \neq \beta_0' x_i + \beta_{10}' z_{1i}$, so $E[\delta(x_i, z_{1i})|z_i] = (\beta_0 - \tilde{\beta})' E[x_i|z_i] + (\beta_{10} - \tilde{\beta}_1)' z_{1i}$. Therefore $E[\delta(x_i, z_{1i})|z_i] = 0$ implies $\delta(x_i, z_{1i}) = 0$ - or $\beta_0 = \tilde{\beta}$ and $\beta_{10} = \tilde{\beta}_1$ - if z_i satisfies the standard rank condition (e.g., it includes excluded instruments from z_{1i} that are correlated with x_i).

For the general non-parametric case, a sufficient condition for identification is that the conditional distribution of x_i given z_i satisfies the completeness condition (see Newey and Powell (2003) or Hall and Horowitz (2005)). The condition implies that $E[\delta(x_i, z_{1i})|z_i] = 0$ implies $\delta(x_i, z_{1i}) = 0$ for any $\delta(x_i, z_{1i})$ with finite expectation. In this sense the completeness condition is nothing but a nonparametric version of the rank condition for identification.

5 Estimation

Our estimator is obtained in three steps. We focus on sieve estimation because it is convenient to impose the restriction (12). We use capital letters to denote random variables and lower case letters to denote their realizations. We assume the tuple $\{(Y_i, X_i, Z_i)\}$ for $i = 1, \ldots, n$ are i.i.d. We let X_i be $d_x \times 1$, Z_{1i} be $d_1 \times 1$, Z_{2i} be $d_2 \times 1$, $d_z = d_1 + d_2$ and $d = d_z + d_x$, with $d_x = 1$ for ease of exposition. Let $\{p_j(Z), j = 1, 2, \ldots\}$ denote a sequence of approximating basis functions (e.g. orthonormal polynomials or splines). Let $p^{k_n} = (p_1(Z), \ldots, p_{k_n}(Z))'$, $P = (p^{k_n}(Z_1), \ldots, p^{k_n}(Z_n))'$, and $(P'P)^-$ denote the Moore-Penrose generalized inverse, where k_n tends to infinity but $k_n/n \to 0$. Similarly we let $\{\phi_j(X, Z_1), j = 1, 2, \ldots\}$ denote a sequence of approximating basis functions, $\phi^{Kn} = (\phi_1(X, Z_1), \ldots, \phi_{K_n}(X, Z_1))'$, where K_n tends to infinity but $K_n/n \to 0$.[9]

In the first step to estimate the controls we estimate $\Pi_0(z)$ using

$$\hat{\Pi}(z) = p^{k_n}(z)'(P'P)^- \sum_{i=1}^{n} p^{k_n}(z_i)x_i$$

and obtain the control variable as $\hat{v} = x - \hat{\Pi}(z)$.

In the second step we construct approximating basis functions using $\hat{v}$ and z, where we impose the CMR condition (12) by subtracting out the conditional means (conditional on Z). We start by assuming v is known and then show how the setup changes when $\hat{v}$ replaces v. We write basis functions when v is known as

$$\tilde{\varphi}_l(z, v) = \varphi_l(z, v) - \bar{\varphi}_l(z)$$

where $\bar{\varphi}_l(z) = E[\varphi_l(V, Z)|Z = z]$ and $\{\varphi_l(z, v), l = 1, 2, \ldots\}$ denotes a sequence of approximating basis functions generated using $(z, v) \in \mathcal{Z} \times \mathcal{V} \equiv \mathcal{W}$, the support of (Z, V). We let $\mathcal{H}$ denote a space of functions that includes h_0, and we let $\|\cdot\|_{\mathcal{H}}$ be a pseudo-metric on $\mathcal{H}$. We define the sieve space $\mathcal{H}_n$ as the collection of functions

$$\mathcal{H}_n = \{h : h = \sum_{l \le L_n} a_l \tilde{\varphi}_l(z, v), \|h\|_{\mathcal{H}} < \bar{C}_h, (z, v) \in \mathcal{W}\}$$

for some bounded positive constant $\bar{C}_h$, with $L_n \to \infty$ so that $\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \ldots \subseteq \mathcal{H}$ (and $L_n/n \to 0$).

Because v is not known we use instead estimates of the approximating basis functions, which we denote as $\hat{\tilde{\varphi}}_l(z, \hat{v}) = \varphi_l(z, \hat{v}) - \hat{\bar{\varphi}}_l(z)$, where $\hat{\bar{\varphi}}_l(z) = \hat{E}[\varphi_l(Z, \hat{V})|Z = z]$. We then construct the approximation of $h(z, v)$ as [10]

$$\begin{aligned} \hat{h}_{L_n}(z, \hat{v}) &= \sum_{l=1}^{Ln} a_l \{\varphi_l(z, \hat{v}) - \hat{E}[\varphi_l(Z, \hat{V})|Z = z]\} \\ &= \sum_{l=1}^{Ln} a_l \{\varphi_l(z, \hat{v}) - p^{k_n}(z)'(P'P)^- \sum_{i=1}^{n} p^{k_n}(z_i)\varphi_l(z_i, \hat{v}_i)\}, \end{aligned} \tag{15}$$

with coefficients, $(a_1, \ldots, a_{L_n})$ to be estimated in the last step. We approximate the sieve

[9] We state specific rate conditions in the next section for our convergence rate results and also for $\sqrt{n}$-consistency and asymptotic normality of linear functionals.

[10] We can use different sieves (e.g., power series, splines of different lengths) to approximate $E[\varphi_l(Z, V)|Z = z]$ and $\Pi(z)$ depending on their smoothness, but we assume one uses the same sieves for notational simplicity.

space $\mathcal{H}_n$ with $\hat{\mathcal{H}}_n$ using (15), so $\hat{\mathcal{H}}_n$ is given by

$$\hat{\mathcal{H}}_n = \{h : h = \sum\nolimits_{l \leq L_n} a_l \hat{\tilde{\varphi}}_l(z, \hat{v}), \|h\|_{\mathcal{H}} < \bar{C}_h, (z, \hat{v}) \in \mathcal{W}\}. \tag{16}$$

In the last step we define $\mathcal{F}$ as the space of functions that includes f_0, and we let $\|\cdot\|_{\mathcal{F}}$ be a pseudo-metric on $\mathcal{F}$. We define the sieve space $\mathcal{F}_n$ as the collection of functions

$$\mathcal{F}_n = \{f : f = \sum\nolimits_{l \leq K_n} \beta_l \phi_l(x, z_1), \|f\|_{\mathcal{F}} < \bar{C}_f, (x, z_1) \in \mathcal{X} \times \mathcal{Z}_1\}$$

for some bounded positive constant $\bar{C}_f$, with $K_n \to \infty$ so that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \ldots \subseteq \mathcal{F}$ (and $K_n/n \to 0$). Then our multi-step series estimator is obtained by solving

$$(\hat{f}, \hat{h}) = \text{arginf}_{(f,h) \in \mathcal{F}_n \times \hat{\mathcal{H}}_n} \sum_{i=1}^{n} \{y_i - (f(x_i, z_{1i}) + h(z_i, \hat{v}_i))\}^2/n \tag{17}$$

where $\hat{v}_i = x_i - \hat{\Pi}(z_i)$.

Equivalently we can write

$$\min_{(\beta_1, \ldots, \beta_{K_n}, a_1, \ldots, a_{L_n})} \sum\nolimits_{i=1}^{n} \{y_i - (\sum\nolimits_{k=1}^{K_n} \beta_k \phi_k(x_i, z_{1i}) + \sum\nolimits_{l=1}^{Ln} a_l \hat{\tilde{\varphi}}_l(z_i, \hat{v}_i))\}^2/n.$$

With fixed k_n, L_n, and K_n our estimator is just a three-stage least squares estimator. Once we obtain the estimates $(\hat{f}, \hat{h})$ we can also estimate linear functionals of (f_0, h_0) using plug-in methods (see Section 7). Next we provide the convergence rates of the nonparametric estimators.

6 Convergence rates

We obtain the convergence rates building on Newey, Powell, and Vella (1999). We differ from their approach as we have another nonparametric estimator in the middle step of estimation. We derive the mean-squared error convergence rates of the nonparametric estimator $\hat{f}(\cdot)$ and $\hat{h}(\cdot)$, which we later use to obtain the $\sqrt{n}$-consistency and the asymptotic normality of the linear functionals of (f_0, h_0).

We introduce additional notation. We let $g_0(z_i, v_i) = f_0(x_i, z_{1i}) + h_0(z_i, v_i)$ be a function of (z_i, v_i) (x_i is fixed given (z_i, v_i)). For a random matrix D, let $\|D\| = (\text{tr}(D'D))^{1/2}$, and let $\|D\|_\infty$ be the infimum of constants C such that $\Pr(||D|| < C) = 1$. Assumptions C1 and C2 together ensure that we obtain the mean-squared error convergence of $\hat{g} = \hat{f} + \hat{h}$ to g_0, and so that of $\hat{f}$ to f_0, too.

Assumption 1 (C1)**.** *(i)* $\{(Y_i, X_i, Z_i)\}_{i=1}^n$ *are i.i.d.,* $V_i = X_i - E[X_i|Z_i]$*, and* $var(X|Z)$, $var(Y|Z, V)$*, and* $var(\varphi_l(Z, V)|Z)$ *for all* l *are bounded; (ii)* (Z, X) *are continuously dis-*

tributed with densities that are bounded away from zero on their supports, which are compact; (iii) $\Pi_0(z)$ *is continuously differentiable of order* s_1 *and all the derivatives of order* s_1 *are bounded on the support of* Z*; (iv)* $\bar{\varphi}_l(Z)$ *is continuously differentiable of order* s_2 *and all the derivatives of order* s_2 *are bounded for all* l *on the support of* Z*; (v)* $h_0(Z,V)$ *is Lipschitz and is continuously differentiable of order* s *and all the derivatives of order* s *are bounded on the support of* (Z,V)*; (vi)* $\varphi_l(z,v)$ *is Lipschitz and is twice continuously differentiable in* v *and its first and second derivatives are bounded for all* l*; (vii)* $f_0(X,Z_1)$ *is continuously differentiable of order* s *and all the derivatives of order* s *are bounded on the support of* (X,Z_1)*.*

Assumptions C1 (iii), (iv), (v), and (vii) ensure that the unknown functions $\Pi_0(Z)$, $\bar{\varphi}_l(Z)$, $h_0(Z,V)$, and $f_0(X,Z_1)$ belong to a Hölder class of functions, so they can be approximated up to the orders of $O(k_n^{-s_1/d_z})$, $O(k_n^{-s_2/d_z})$, $O(L_n^{-s/d})$, and $O(K_n^{-s/(d_x+d_1)})$ respectively when using polynomials or splines (see Timan (1963), Schumaker (1981), Newey (1997), and Chen (2007)). Assumption C1 (vi) is satisfied for polynomial and spline basis functions with appropriate orders. Assumption C1 (ii) can be relaxed with some additional complexity (e.g., a trimming device as in Newey, Powell, and Vella (1999)). Assumption C1 (v) and (vii) maintain that f_0 and h_0 have the same order of smoothness for ease of notation, but it is possible to allow them to differ.

Next we impose the rate conditions that restrict the growth of k_n, K_n, and L_n as n tends to infinity. We write $\mathbf{L}_n = K_n + L_n$.

Assumption 2 (C2)**.** *Let* $\triangle_{n,1} = k_n^{1/2}/\sqrt{n} + k_n^{-s_1/d_z}$, $\triangle_{n,2} = k_n^{1/2}/\sqrt{n} + k_n^{-s_2/d_z}$*, and* $\triangle_n = \max\{\triangle_{n,1}, \triangle_{n,2}\}$*. For polynomial approximations* $\mathbf{L}_n^{1/2}(L_n^3 + L_n^{1/2}k_n^{3/2}/\sqrt{n} + L_n^{1/2})\triangle_n \to 0$, $\mathbf{L}_n^3/n \to 0$*, and* $k_n^3/n \to 0$*. For spline approximations* $\mathbf{L}_n^{1/2}(L_n^{3/2} + L_n^{1/2}k_n/\sqrt{n} + L_n^{1/2})\triangle_n \to 0$ *,* $\mathbf{L}_n^2/n \to 0$*, and* $k_n^2/n \to 0$*.*

Theorem 5. *Suppose Assumptions C1-C2 are satisfied. Then*

$$\left(\int (\hat{g}(z,v) - g(z,v))^2 d\mu_0(z,v)\right)^{1/2} = O_p(\sqrt{\mathbf{L}_n/n} + L_n\triangle_n + \mathbf{L}_n^{-s/d}).$$

where $\mu_0(z,v)$ *denotes the distribution function of* (z,v)*.*

In Theorem 5 the term $L_n\triangle_n$ arises because of the estimation error from the first and second steps of estimation. With no estimation error from these stages we would obtain the convergence rate of $O_p(\sqrt{\mathbf{L}_n/n} + \mathbf{L}_n^{-s/d})$, which is a standard convergence rate of series estimators.

7 Asymptotic Normality

Following Newey (1997) and Newey, Powell, and Vella (1999) we consider inference for the linear functions of g, $\theta = \alpha(g)$. The estimator $\hat{\theta} = \alpha(\hat{g})$ of $\theta_0 = \alpha(g_0)$ is a well-defined "plug-in" estimator, and because of the linearity of $\alpha(g)$ we have

$$\hat{\theta} = \mathcal{A}\hat{\beta}, \mathcal{A} = (\alpha(\phi_1), \ldots, \alpha(\phi_{K_n}), \alpha(\tilde{\varphi}_1), \ldots, \alpha(\tilde{\varphi}_{L_n}))$$

where we let $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_{K_n}, \hat{a}_1, \ldots, \hat{a}_{L_n})'$. This setup includes (e.g.) partially linear models, where f contains some parametric components, and the weighted average derivative, where one estimates the average response of y with respect to the marginal change of x or z_1. More generally, if $\mathcal{A}$ depends on unknown population objects, we can estimate it using $\hat{\mathcal{A}} = \partial\alpha(\hat{\hat{\psi}}_i^{\mathbf{L}\prime}\beta)/\partial\beta'|_{\beta=\hat{\beta}}$ where $\hat{\hat{\psi}}_i^{\mathbf{L}} = (\phi_1(x_i, z_{1i}), \ldots, \phi_K(x_i, z_{1i}), \hat{\bar{\varphi}}_1(z_i, \hat{v}_i), \ldots, \hat{\bar{\varphi}}_L(z_i, \hat{v}_i))'$, so that $\hat{\theta} = \hat{\mathcal{A}}\hat{\beta}$ (see Newey (1997)).

We focus on conditions that provide for $\sqrt{n}$-asymptotics and allow for a straightforward consistent estimator for the standard errors of $\hat{\theta}$.[11] If there exists a Riesz representer $\nu^*(Z, V)$ such that

$$\alpha(g) = E[\nu^*(Z, V)g(Z, V)] \tag{18}$$

for any $g = (f, h) \in \mathcal{F} \times \mathcal{H}$ that can be approximated by power series or splines in the mean-squared norm, then we can obtain $\sqrt{n}$-consistency and asymptotic normality for $\hat{\theta}$, expressed as

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \Omega),$$

for some asymptotic variance matrix Ω. In Assumption C1 we take both $\mathcal{F}$ and $\mathcal{H}$ as Hölder spaces of functions, which ensures the approximation of g in the mean-squared norm (see e.g., Newey (1997), Newey, Powell, and Vella (1999), and Chen (2007)). Letting $\rho_v(Z) = E[\nu^*(Z, V)(\frac{\partial h_0(Z,V)}{\partial V} - E[\frac{\partial h_0(Z,V)}{\partial V}|Z])|Z]$ and $\rho_{\bar{\varphi}_l}(Z) = E[a_l\nu^*(Z, V)|Z]$, the asymptotic variance of the estimator $\hat{\theta}$ is given by

$$\begin{aligned}\Omega &= E[\nu^*(Z, V)\text{var}(Y|Z, V)\nu^*(Z, V)'] + E[\rho_v(Z)\text{var}(X|Z)\rho_v(Z)'] \\ &\quad + \lim_{n\to\infty}\sum_{l=1}^{L_n} E[\rho_{\bar{\varphi}_l}(Z)\text{var}(\varphi_l(Z, V)|Z)\rho_{\bar{\varphi}_l}(Z)'].\end{aligned} \tag{19}$$

The first term in the variance accounts for the final stage of estimation, the second term accounts for the estimation of the control (v), and the last term accounts for the middle step of the estimation.

Assumption C1, R1, N1, and N2 below are sufficient for us to characterize the asymp-

[11] Developing the asymptotic distributions of the functionals that do not yield the $\sqrt{n}$-consistency is also possible based on the convergence rates result we obtained and alternative assumptions on the functionals of interest (see Newey, Powell, and Vella (1999)).

totic normality of $\hat{\theta}$ and also a consistent estimator for the asymptotic variance of $\hat{\theta}$. Let $\psi^{\mathbf{L}}(z_i, v_i) \equiv (\phi_1(x_i, z_{1i}), \ldots, \phi_K(x_i, z_{1i}), \tilde{\varphi}^L(z_i, v_i)')'$ and $\tilde{\varphi}^L(z_i, v_i) = (\tilde{\varphi}_1(z_i, v_i), \ldots, \tilde{\varphi}_L(z_i, v_i))'$.

Assumption 3 (R1)**.** *There exist $\nu^*(Z, V)$ and $\beta_{\mathbf{L}}$ such that $E[||\nu^*(Z, V)||^2] < \infty$, $\alpha(g_0) = E[\nu^*(Z, V)g_0(Z, V)]$, $\alpha(\phi_k) = E[\nu^*(Z, V)\phi_k]$ for $k = 1, \ldots, K$, $\alpha(\tilde{\varphi}_l) = E[\nu^*(Z, V)\tilde{\varphi}_l]$ for $l = 1, \ldots, L$, and $E[||\nu^*(Z, V) - \psi^{\mathbf{L}}(Z, V)'\beta_{\mathbf{L}}||^2] \to 0$ as $\mathbf{L} \to \infty$.*

To present the theorem, we need additional notation and assumptions. Let $a_L = (a_1, \ldots, a_L)'$ with an abuse of notation and for any differentiable function $c(w)$, let $|\mu| = \sum_{j=1}^{\dim(w)} \mu_j$ and define $\partial^\mu c(w) = \partial^{|\mu|} c(w)/\partial w_1 \cdots \partial w_{\dim(w)}$. Also define $|c(w)|_\delta = \max_{|\mu| \leq \delta} \sup_{w \in \mathcal{W}} ||\partial^\mu c(w)||$ and others are defined similarly.

Assumption 4 (N1)**.** *(i) there exist δ, γ, and $\beta_{\mathbf{L}}$ such that $|g_0(z, v) - \beta'_{\mathbf{L}}\psi^{\mathbf{L}}(z, v)|_\delta \leq C\mathbf{L}^{-\gamma}$ (which also implies $|h_0(z, v) - a'_L\tilde{\varphi}^L(z, v)|_\delta \leq CL^{-\gamma}$); (ii) $var(Y_i|Z_i, V_i)$ is bounded away from zero, $E[\eta_i^4|Z_i, V_i]$ and $E[V_i^4|Z_i]$ are bounded and $E[\tilde{\varphi}_l(Z_i, V_i)^4|Z_i]$ is bounded for all l.*

The assumption N1 (i) is satisfied for f_0 and h_0 that belong to the Hölder class. Then we can take (e.g.) $\gamma = s/d$. Next we impose the rate conditions that restrict the growth of k_n and $\mathbf{L}_n = K_n + L_n$ as n tends to infinity.

Assumption 5 (N2)**.** *Let $\triangle_{n,1} = k_n^{1/2}/\sqrt{n} + k_n^{-s_1/d_z}$, $\triangle_{n,2} = k_n^{1/2}/\sqrt{n} + k_n^{-s_2/d_z}$, and $\triangle_n = \max\{\triangle_{n,1}, \triangle_{n,2}\}$. $\sqrt{n}k_n^{-s_1/d_z} \to 0$, $\sqrt{n}k_n^{-s_2/d_z} \to 0$, $\sqrt{n}k_n^{1/2}L_n^{-s/d} \to 0$, $\sqrt{n}\mathbf{L}_n^{-s/d} \to 0$ and they are sufficiently small. For the polynomial approximations $\frac{\mathbf{L}_n^2 + \mathbf{L}_n L_n^3 k_n + \mathbf{L}_n^{1/2}(L_n^4 k_n^{3/2} + k_n^{5/2})}{\sqrt{n}} \to 0$ and for the spline approximations $\frac{\mathbf{L}_n^{3/2} + \mathbf{L}_n L_n^{3/2} k_n^{1/2} + \mathbf{L}_n^{1/2}(L_n^{5/2} k_n + k_n^{3/2}) + L_n^{3/2} k_n^{3/2}}{\sqrt{n}} \to 0$.*

Theorem 6. *Suppose Assumptions C1, R1, and N1-N2 are satisfied. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to_d N(0, \Omega).$$

Based on this asymptotic distribution, one can construct the confidence intervals of θ_0 and calculate standard errors in a straightforward manner. Let $\hat{g}(z_i, \hat{v}_i) = \hat{f}(x_i, z_{1i}) + \hat{h}(z_i, \hat{v}_i)$ and $\hat{g}_i = \hat{g}(z_i, \hat{v}_i)$. Define $\hat{\tilde{\psi}}_i^{\mathbf{L}} = (\phi_1(x_i, z_{1i}), \ldots, \phi_K(x_i, z_{1i}), \hat{\tilde{\varphi}}^L(z_i, \hat{v}_i)')'$ where $\hat{\tilde{\varphi}}^L(z_i, v_i) = (\hat{\tilde{\varphi}}_1(z_i, v_i), \ldots, \hat{\tilde{\varphi}}_L(z_i, v_i))'$. Let

$$\begin{aligned}
\hat{\mathcal{T}} &= \sum\nolimits_{i=1}^{n} \hat{\hat{\psi}}_i^{\mathbf{L}} \hat{\hat{\psi}}_i^{\mathbf{L}\prime}/n, \hat{\Sigma} = \sum\nolimits_{i=1}^{n} (y_i - \hat{g}(z_i, \hat{v}_i))^2 \hat{\hat{\psi}}_i^{\mathbf{L}} \hat{\hat{\psi}}_i^{\mathbf{L}\prime}/n \qquad (20) \\
\hat{\mathcal{T}}_1 &= P'P/n, \hat{\Sigma}_1 = \sum_{i=1}^{n} \hat{v}_i^2 p^k(z_i) p^k(z_i)'/n, \hat{\Sigma}_{2,l} = \sum_{i=1}^{n} \{\varphi_l(z_i, \hat{v}_i) - \hat{\bar{\varphi}}_l(z_i)\}^2 p^k(z_i) p^k(z_i)'/n \\
\hat{H}_{11} &= \sum_{i=1}^{n} \sum_{l=1}^{L} \hat{a}_l \frac{\partial \varphi_l(z_i, \hat{v}_i)}{\partial v_i} \hat{\hat{\psi}}_i^{\mathbf{L}} p^k(z_i)'/n, \\
\hat{H}_{12} &= \sum_{i=1}^{n} p^k(z_i)'((P'P)^{-} \sum_{j=1}^{n} p^k(z_j) \frac{\partial \sum_{l=1}^{L} \hat{a}_l \varphi_l(z_j, \hat{v}_j)}{\partial v_j}) \hat{\hat{\psi}}_i^{\mathbf{L}} p^k(z_i)'/n, \\
\hat{H}_{2,l} &= \sum\nolimits_{i=1}^{n} \hat{a}_l \hat{\hat{\psi}}_i^{\mathbf{L}} p^k(z_i)'/n, \hat{H}_1 = \hat{H}_{11} - \hat{H}_{12}.
\end{aligned}$$

Then, we can estimate Ω consistently by

$$\hat{\Omega} = \mathcal{A}\hat{\mathcal{T}}^{-1} \left[\hat{\Sigma} + \hat{H}_1 \hat{\mathcal{T}}_1^{-1} \hat{\Sigma}_1 \hat{\mathcal{T}}_1^{-1} \hat{H}_1' + \sum\nolimits_{l=1}^{L_n} \hat{H}_{2,l} \hat{\mathcal{T}}_1^{-1} \hat{\Sigma}_{2,l} \hat{\mathcal{T}}_1^{-1} \hat{H}_{2,l}' \right] \hat{\mathcal{T}}^{-1} \mathcal{A}'. \qquad (21)$$

Theorem 7. *Suppose Assumptions C1, R1, and N1-N2 are satisfied. Then* $\hat{\Omega} \to_p \Omega$.

This is the heteroskedasticity robust variance estimator that accounts for the first and second steps of estimation. The first variance term $\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\Sigma}\hat{\mathcal{T}}^{-1}\mathcal{A}'$ corresponds to the variance estimator without error from the first and second steps of estimation. The second variance term accounts for the estimation of v (and corresponds to the second term in (19)). The third variance term accounts for the estimation of $\bar{\varphi}_l(\cdot)$'s) (and corresponds to the third term in (19)). If we view our model as a parametric one with fixed k_n, K_n, and L_n, the same variance estimator $\hat{\Omega}$ can be used as the estimator of the variance for the parametric model (e.g, Newey (1984) and Murphy and Topel (1985)).

7.1 Discussion

We discuss Assumption R1 for the partially linear model and the weighted average derivative. Consider a partially linear model of the form

$$f_0(x, z_1) = x_1'\beta_{10} + f_{20}(x_{-1}, z_1)$$

where x can be multi-dimensional and x_1 is a subvector of x such that $x = (x_1, x_{-1})$. Then we have

$$\beta_{10} = \alpha(g_0) = E[\nu^*(Z, V) g_0(Z, V)]$$

where $\nu^*(z, v) = (E[q(Z, V)q(Z, V)'])^{-1} q(z, v)$ and $q(z, v)$ is the residual from the mean-square projection of x_1 on the space of functions that are additive in (x_{-1}, z_1) and any $h(z, v)$ such that $E[h(Z, V)|Z] = 0$.[12] Thus we can approximate $q(z, v)$ by the mean-square

[12]Note that existence of the Riesz representer in this setting requires $E[q(Z, V)q(Z, V)']$ to be nonsingular.

projection residual of x_1 on $\psi^{\mathbf{L}}_{-1}(z_i, v_i) \equiv (\phi_1(x_{-1i}, z_{1i}), \ldots, \phi_K(x_{-1i}, z_{1i}), \tilde{\varphi}^L(z_i, v_i)')'$, and then use these estimates to approximate $\nu^*(z, v)$.

Next consider a weighted average derivative of the form

$$\alpha(g_0) = \int_{\bar{\mathcal{W}}} \varpi(x, z_1, \kappa(z, v)) \frac{\partial g_0(z, v)}{\partial x} d(z, v) = \int \varpi(x, z_1, \kappa(z, v)) \frac{\partial f_0(x, z_1)}{\partial x} d(z, v)$$

where the weight function $\varpi(x, z_1, \kappa(z, v))$ puts zero weights outside $\bar{\mathcal{W}} \subset \mathcal{W}$ and $\kappa(z, v)$ is some function such that $E[\kappa(Z, V)|Z] = 0$. This is a linear functional of g_0. Integration by parts shows that

$$\alpha(g_0) = -\int_{\bar{\mathcal{W}}} \text{proj}(\mu_0(z, v)^{-1} \frac{\partial \varpi(x, z_1, \kappa(z, v))}{\partial x} |\mathcal{S}) g_0(z, v) d\mu_0(z, v) = E[\nu^*(Z, V) g(Z, V)]$$

where $\text{proj}(\cdot|\mathcal{S})$ denotes the mean-square projection on the space of functions that are additive in (x, z_1) and any $h(z, v)$ such that $E[h(Z, V)|Z] = 0$ (so the Riesz representer $\nu^*(z, v)$ is well-defined), and $\nu^*(z, v) = -\text{proj}(\mu_0(z, v)^{-1} \frac{\partial \varpi(x, z_1, \kappa(z, v))}{\partial x} |\mathcal{S})$ with $\mu_0(z, v)$ denoting the distribution of (z, v). We can then approximate $\nu^*(z, v)$ using a mean-square projection of $\mu_0(z, v)^{-1} \frac{\partial \varpi(x, z_1, \kappa(z, v))}{\partial x}$ on $\psi^{\mathbf{L}}(z_i, v_i)$.

8 Simulation Study

We investigate the performance of our **CMRCF** estimator in nonlinear additive models when the classic CF condition (3) does not hold. We compare three estimators: the classical CF estimator that assumes $E[\varepsilon_i|z_i, v_i] = E[\varepsilon_i|v_i] = \rho v_i$, a simple version of NPVCF estimator that maintains $E[\varepsilon_i|z_i, v_i] = E[\varepsilon_i|v_i]$, and our new CMRCF estimator. Our simulation results illustrate that maintaining the classic CF restriction can produce biased estimates while the **CMRCF** estimator performs well when the classic CF condition does not hold.

We consider the following simultaneous equations models :

$$\begin{aligned}
[1]\ y_i &= \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i) \cdot \log(z_i) \\
[2]\ y_i &= \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i) / \exp(z_i) \\
[3]\ y_i &= \alpha + \beta x_i + \gamma \log x_i + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i) / \exp(z_i) \\
[4]\ y_i &= \alpha + \beta x_i + \gamma \log x_i + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i + \varepsilon_i \cdot \varsigma_i) / \exp(z_i) \\
[5]\ y_i &= \alpha + \beta x_i + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i) / \exp(z_i) \\
[6]\ y_i &= \alpha + \beta x_i + \gamma x_i^2 + \varepsilon_i \ ; \ x_i = z_i + (3\varepsilon_i + \varsigma_i).
\end{aligned}$$

In all designs [1]-[6], x_i is correlated with ε_i, and the CMR condition holds.[13] Except for

[13]For example, in design [2] $E[\varepsilon_i|z_i] = E[E[\varepsilon_i|z_i, v_i]|z_i] = \frac{\exp(z_i)E[v_i|z_i] - E[\varsigma_i|z_i]}{3} = 0$ since $E[v_i|z_i] = 0$ and

design [6], the classic **CF** restriction does not hold.[14] We generate simulation data based on the following distributions: $\varepsilon_i \sim U_{[-1/2,1/2]}$, $\varsigma_i \sim U_{[-1/2,1/2]}$, $z_i = 2 + 2U_{[-1/2,1/2]}$, where $U_{[-1/2,1/2]}$ denotes the uniform distribution supported on $[-1/2, 1/2]$, and all three random variables are independent of one another. We set the true parameter values $(\alpha_0, \beta_0, \gamma_0) = (1, 1, -1)$. The data is generated with the sample sizes: $n = 1,000$.

The classic CF (CCF) estimates

$$y_i = f(x_i) + \rho\hat{v}_i + \eta_i$$

using the first stage estimation residual $\hat{v}_i = x_i - (\hat{\pi}_0 + \hat{\pi}_1 z_i + \hat{\pi}_2 z_i^2)$ where $f(x_i)$ is given by the designs [1]-[6]. The NPVCF estimator is obtained by estimating

$$y_i = f(x_i) + h(\hat{v}_i) + \eta_i,$$

where we approximate $h(\hat{v}_i)$ as $h(\hat{v}_i) = \sum_{l=1}^{5} a_l \hat{v}_i^l$.[15] Since the NPVCF does not identify the constant term by design, we normalize $h(\hat{v}_i) = 0$ in the estimation so that the constant term (α) is correctly obtained.

We obtain the CMRCF estimator as follows. Using the first stage estimation residual $\hat{v}_i$ construct approximating functions $\tilde{v}_{1i} = \hat{v}_i$, $\tilde{v}_{2i} = \hat{v}_i^2 - \hat{E}[\hat{v}_i^2|z_i]$, $\tilde{v}_{3i} = \hat{v}_i^3 - \hat{E}[\hat{v}_i^3|z_i]$, and others (e.g., interactions with polynomials of z_i as $z_i\hat{v}_i$ and $z_i^2\hat{v}_i$) are defined similarly where $\hat{E}[\cdot|z_i]$ is implemented by the least squares estimation on $(1, z_i, z_i^2)$. In the last step we estimate the model parameters using the least squares as described in Section 3. In this simulation design, we estimate the parameters as

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{a}) = \text{argmin} \sum_{i=1}^{n} \{y_i - (f(x_i; \alpha, \beta, \gamma) + h(z_i, \hat{v}_i))\}^2/n$$

where $h(z_i, \hat{v}_i) = \sum_{l=1}^{L} a_l \tilde{v}_{li}$ depends on the simulation designs. The choice of the basis in the finite sample is not a consistency issue but it is an efficiency issue, and we vary this choice across specifications. In design [1] we use $\tilde{v}_{1i}$ and $z_i\tilde{v}_i$ as the controls. In designs [2], [5], and [6] we use the controls $\tilde{v}_{1i}, \tilde{v}_{2i}$, and $z_i\tilde{v}_i$. In design [3] we use the controls $\tilde{v}_{1i}, \tilde{v}_{2i}, z_i\tilde{v}_i$, and $z_i^2\tilde{v}_i$, and in design [4] we use $\tilde{v}_{1i}, \tilde{v}_{2i}, \tilde{v}_{3i}, \tilde{v}_{4i}, z_i\tilde{v}_i$.

$E[\varsigma_i|z_i] = 0$ by construction.

[14] For example, in design [2] we have $v_i = x_i - E[x_i|z_i] = (3\varepsilon_i + \varsigma_i)/\exp(z_i)$ (we will let the error terms ε_i and ς_i be independent with the instrument z_i). Then we have $\varepsilon_i = (\exp(z_i)v_i - \varsigma_i)/3$ and therefore $E[\varepsilon_i|z_i, v_i] = (\exp(z_i)v_i - E[\varsigma_i|z_i, v_i])/3$, and this cannot be written as a function of v_i only.

[15] We do not use the trimming device in Newey, Powell, and Vella (1999). The trimming is not important here because the supports of variables are compact and tightly bounded.

We report the biases and the RMSE's based on 200 repetitions of the estimations. The simulation results in Tables I-VI show that CCF and NPVCF are biased in all designs except [5] and [6], for which the theory says they should be consistent. The CMRCF is robust regardless of the designs. In design [5] all three approaches produce correct estimates because the outcome equation is linear, which is consistent with our discussion in Section 2. In design [6] all three approaches are consistent because the classic **CF** restriction holds. From these simulation results we conclude that our CMRCF approach performs well regardless of whether the model is linear or nonlinear or whether the classic **CF** restriction holds (or not), while the CCF and NPVCF approaches critically hinge on the model and the **CF** restriction.

9 Conclusion

The problem of endogenous regressors in simultaneous equations models has a long history in econometrics and empirical studies. In models with additively separable errors researchers have used both the instrumental variable (IV) approach and the classical control function (CF) approach to correct for the bias induced by the correlation between the error and the regressor(s). We revisit the well-known numerical equivalence result between two-stage least squares (2SLS) and the classic CF. We show this equivalence raises an interesting but unrecognized puzzle. The classic CF approach maintains that the regression error is mean independent of the instruments conditional on the CF control, which is not required by 2SLS, and could easily be violated.

We show that the classic CF estimator can be modified to allow the mean of the error to depend in a general way on the instruments and control. We do so by replacing the classic CF restriction with a generalized CF moment condition combined with the moment restrictions maintained by 2SLS or nonparametric 2SLS. We then show that our generalized CF estimator is consistent in parametric or non-parametric settings with endogenous regressors and additive errors.

If the outcome equation is nonlinear/nonparametric in the endogenous regressor, then both the classical CF estimator and the NPVCF estimator of Newey, Powell, and Vella (1999) are inconsistent when the classic CF restriction does not hold while our estimator remains consistent. Our new approach is not subject to the ill-posed inverse problem and can be estimated using a multi-step least squares in the simplest case. Therefore our new estimator combines the strengths of both the NP2SLS estimator of Newey and Powell (2003) and the NPVCF estimator of Newey, Powell, and Vella (1999) while avoiding a weakness of each of these approaches. Our simulation study shows that the classic CF estimator and the NPVCF estimator are biased when the CF restriction is violated while our estimator remains consistent.

Table I: Design [1], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = -1$

Nonlinear & **CF** condition does not hold

		mean	bias	RMSE
CCF	α	0.7076	-0.2924	0.2952
	β	1.3078	0.3078	0.3094
	γ	-1.0679	-0.0679	0.0682
NPVCF	α	0.7010	-0.2990	0.3034
	β	1.3677	0.3677	0.3738
	γ	-1.0917	-0.0917	0.0938
CMRCF	α	0.9978	-0.0022	0.0548
	β	1.0021	0.0021	0.0503
	γ	-1.0005	-0.0005	0.0109

Table II: Design [2], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = -1$

Nonlinear & **CF** condition does not hold

		mean	bias	RMSE
CCF	α	1.5331	0.5531	0.5452
	β	0.4056	-0.5944	0.6055
	γ	-0.8496	0.1504	0.1529
NPVCF	α	1.3466	0.3466	0.3697
	β	0.6283	-0.3717	0.3948
	γ	-0.9090	0.0910	0.0966
CMRCF	α	0.9933	-0.0067	0.1478
	β	1.0079	0.0079	0.1611
	γ	-1.0021	-0.0021	0.0405

Table III: Design [3], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = -1$

Nonlinear & **CF** condition does not hold

		mean	bias	RMSE
CCF	α	0.5818	-0.4182	0.4235
	β	1.5048	0.5048	0.5108
	γ	-1.9246	-0.9246	0.9367
NPVCF	α	0.7666	-0.2334	0.2482
	β	1.3042	0.3042	0.3200
	γ	-1.5861	-0.5861	0.6156
CMRCF	α	0.9943	-0.0057	0.1103
	β	1.0076	0.0076	0.1255
	γ	-1.0144	-0.0144	0.2249

Table IV: Design [4], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = -1$

Nonlinear & **CF** condition does not hold

		mean	bias	RMSE
CCF	α	0.6109	-0.3891	0.3950
	β	1.4702	0.4702	0.4769
	γ	-1.8617	-0.8617	0.8751
NPVCF	α	0.7614	-0.2386	0.2541
	β	1.3333	0.3333	0.3497
	γ	-1.6687	-0.6687	0.6988
CMRCF	α	1.0003	0.0003	0.1117
	β	1.0005	0.0005	0.1267
	γ	-1.0016	-0.0016	0.2262

Table V: Design [5], $\alpha_0 = 1, \beta_0 = 1$

Linear & **CF** condition does not hold

		mean	bias	RMSE
CCF	α	0.9993	-0.0007	0.0343
	β	1.0004	0.0004	0.0172
		.	.	.
NPVCF	α	1.0007	0.0007	0.0384
	β	0.9997	-0.0003	0.0192
		.	.	.
CMRCF	α	0.9991	-0.0009	0.0343
	β	1.0005	0.0005	0.0171
		.	.	.

Table VI: Design [6], $\alpha_0 = 1, \beta_0 = 1, \gamma_0 = -1$

Nonlinear & **CF** condition holds

		mean	bias	RMSE
CCF	α	0.9991	-0.0009	0.0354
	β	1.0010	0.0010	0.0200
	γ	-1.0002	-0.0002	0.0024
NPVCF	α	0.9997	-0.0003	0.0338
	β	1.0004	0.0004	0.0210
	γ	-1.0001	-0.0001	0.0032
CMRCF	α	0.9975	-0.0025	0.0891
	β	1.0068	0.0068	0.1204
	γ	-1.0021	-0.0021	0.0304

Appendix

A Proof of convergence rates

We first introduce notation and prove Lemma L1 below that is useful to prove the convergence rate results.

Define $h_L(z,v) = a_L'\tilde{\varphi}^L(z,v)$ and $\hat{h}_L(z,v) = a_L'\hat{\tilde{\varphi}}^L(z,v)$ where a_L[16] satisfies Assumption L1 (iv). Define $\psi_i^{\mathbf{L}}(z_i,v_i) = (\phi_1(x_i,z_{1i}),\ldots,\phi_K(x_i,z_{1i}),\tilde{\varphi}^L(z_i,v_i)')'$ where $\tilde{\varphi}^L(z_i,v_i) = (\tilde{\varphi}_1(z_i,v_i),\ldots,\tilde{\varphi}_L(z_i,v_i))'$ and $\hat{\psi}_i^{\mathbf{L}}(z_i,v_i) = (\phi_1(x_i,z_{1i}),\ldots,\phi_K(x_i,z_{1i}),\hat{\tilde{\varphi}}^L(z_i,v_i)')'$ with $\hat{\tilde{\varphi}}^L(z_i,v_i) = (\hat{\tilde{\varphi}}_1(z_i,v_i),\ldots,\hat{\tilde{\varphi}}_L(z_i,v_i))'$. We further let $\hat{\hat{\psi}}_i^{\mathbf{L}} = \hat{\psi}^{\mathbf{L}}(z_i,\hat{v}_i)$, $\psi_i^{\mathbf{L}} = \psi^{\mathbf{L}}(z_i,v_i)$, and $\hat{\psi}_i^{\mathbf{L}} = \hat{\psi}^{\mathbf{L}}(z_i,v_i)$. We further let $\psi^{\mathbf{L},n} = (\psi_1^{\mathbf{L}},\ldots,\psi_n^{\mathbf{L}})'$, $\hat{\psi}^{\mathbf{L},n} = (\hat{\psi}_1^{\mathbf{L}},\ldots,\hat{\psi}_n^{\mathbf{L}})'$, and $\hat{\hat{\psi}}^{\mathbf{L},n} = (\hat{\hat{\psi}}_1^{\mathbf{L}},\ldots,\hat{\hat{\psi}}_n^{\mathbf{L}})'$.

Let C (also C_1,C_2, and others) denote a generic positive constant and let $C(Z,V)$ or $C(X,Z_1)$ (also $C_1(\cdot)$, $C_2(\cdot)$, and others) denote a generic bounded positive function of (Z,V) or (X,Z_1). We often write $C_i = C(x_i,z_{1i})$. Recall $\mathcal{W} = \mathcal{Z}\times\mathcal{V}$.

Assumption 6 (L1)**.** *(i) (X,Z,V) is continuously distributed with bounded density; (ii) for each k, L, and $\mathbf{L} = K+L$ there are nonsingular matrices B_1, B_2, and B such that for $p_{B_1}^k(z) = B_1p^k(z)$, $\tilde{\varphi}_{B_2}^L(z,v) = B_2\tilde{\varphi}^L(z,v)$, and $\psi_B^{\mathbf{L}}(z,v) = B\psi^{\mathbf{L}}(z,v)$, $E[p_{B_1}^k(Z_i)p_{B_1}^k(Z_i)']$, $E[\tilde{\varphi}_{B_2}^L(Z_i,V_i)\tilde{\varphi}_{B_2}^L(Z_i,V_i)']$, and $E[\psi_B^{\mathbf{L}}(Z_i,V_i)\psi_B^{\mathbf{L}}(Z_i,V_i)']$ have smallest eigenvalues that are bounded away from zero, uniformly in k, L, and $\mathbf{L}$; (iii) for each integer $\delta>0$, there are $\zeta_\delta(\mathbf{L})$ and $\xi_\delta(k)$ such that $|\psi^{\mathbf{L}}(z,v)|_\delta \le \zeta_\delta(\mathbf{L})$ (this also implies that $|\tilde{\varphi}^L(z,v)|_\delta \le \zeta_\delta(L)$) and $|p^k(z)|_\delta \le \xi_\delta(k)$; (iv) There exist γ, γ_1, $\gamma_2>0$, and $\beta_{\mathbf{L}}$, a_L, λ_k^1, and $\lambda_{l,k}^2$ such that $|\Pi_0(z)-\lambda_k^{1\prime}p^k(z)|_\delta \le Ck^{-\gamma_1}$, $|\bar{\varphi}_{0l}(z)-\lambda_{l,k}^{2\prime}p^k(z)|_\delta \le Ck^{-\gamma_2}$ for all l, $|h_0(z,v)-a_L'\tilde{\varphi}^L(z,v)|_\delta \le CL^{-\gamma}$, and $|g_0(z,v)-\beta_{\mathbf{L}}'\psi^{\mathbf{L}}(z,v)|_\delta \le C\mathbf{L}^{-\gamma}$; (v) both $\mathcal{Z}$ and $\mathcal{X}$ are compact.*

Let $\triangle_{n,1} = k_n^{1/2}/\sqrt{n}+k_n^{-\alpha_1}$ and $\triangle_{n,2} = k_n^{1/2}/\sqrt{n}+k_n^{-\alpha_2}$ and $\triangle_n = \max\{\triangle_{n,1},\triangle_{n,2}\}$.

Lemma 1 (L1)**.** *Suppose Assumptions L1 and Assumptions C1 (i), (vi), (v), (vi), and (vii) hold. Further suppose $\mathbf{L}^{1/2}(\zeta_1(L)+L^{1/2}\xi_0(k)\sqrt{k/n}+L^{1/2})\triangle_n \to 0$, $\xi_0(k)^2k/n\to 0$, and $\zeta_0(\mathbf{L})^2\mathbf{L}/n\to 0$. Then,*

$$\left(\sum_{i=1}^{n}\left(\hat{g}(z_i,v_i)-g_0(z_i,v_i)\right)^2/n\right)^{1/2} = O_p\left(\sqrt{\mathbf{L}/n}+L\xi_0(k)\triangle_{n,1}\sqrt{k/n}+L\triangle_{n,2}+\mathbf{L}^{-\gamma}\right).$$

[16] With abuse of notation we write $a_L = (a_1,\ldots,a_L)'$.

A.1 Proof of Lemma L1

Without loss of generality, we will let $p^k(z) = p^k_{B_1}(z)$, $\tilde{\varphi}^L(z,v) = \tilde{\varphi}^L_{B_2}(z,v)$, and $\psi^{\mathbf{L}}(z,v) = \psi^{\mathbf{L}}_B(z,v)$. Let $\hat{\Pi}_i = \hat{\Pi}(z_i)$ and $\Pi_i = \Pi_0(z_i)$. Let $\hat{\bar{\varphi}}_{l,i} = \hat{\bar{\varphi}}_l(z_i)$ and $\bar{\varphi}_{l,i} = \bar{\varphi}_l(z_i)$. Let $\hat{\tilde{\varphi}}_{l,i} = \hat{\tilde{\varphi}}_l(z_i,\hat{v}_i)$ and $\tilde{\varphi}_{l,i} = \tilde{\varphi}_l(z_i,v_i)$. Also let $\hat{\tilde{\varphi}}^L_i = \hat{\tilde{\varphi}}^L(z_i,\hat{v}_i)$ and $\tilde{\varphi}^L_i = \tilde{\varphi}^L(z_i,v_i)$. Further define $\dot{\bar{\varphi}}_l(z) = p^k(z)'(P'P)^- \sum_{i=1}^n p^k(z_i)\varphi_l(z_i,v_i)$ where we have $\hat{\bar{\varphi}}_l(z) = p^k(z)'(P'P)^- \sum_{i=1}^n p^k(z_i)\varphi_l(z_i,\hat{v}_i)$. Let $\dot{\bar{\varphi}}^L(z) = (\dot{\bar{\varphi}}_1(z),\ldots,\dot{\bar{\varphi}}_L(z))'$ and $\bar{\varphi}^L(z) = (\bar{\varphi}_1(z),\ldots,\bar{\varphi}_L(z))'$. We also let $\varphi^L(z_i,\hat{v}_i) = (\varphi_1(z_i,\hat{v}_i),\ldots,\varphi_L(z_i,\hat{v}_i))'$ and $\varphi^L(z_i,v_i) = (\varphi_1(z_i,v_i),\ldots,\varphi_L(z_i,v_i))'$.

First note $(P'P)/n$ becomes nonsingular w.p.a.1 as $\xi_0(k)^2k/n \to 0$ by Assumption L1 (ii) and the same proof in Theorem 1 of Newey (1997). Then by the same proof (A.3) of Lemma A1 in Newey, Powell, and Vella (1999), we obtain

$$\sum_{i=1}^n ||\hat{\Pi}_i - \Pi_i||^2/n = O_p(\triangle^2_{n,1}) \text{ and } \sum_{i=1}^n ||\dot{\bar{\varphi}}_{l,i} - \bar{\varphi}_{l,i}||^2/n = O_p(\triangle^2_{n,2}) \text{ for all } l. \tag{22}$$

Also by Theorem 1 of Newey (1997), it follows that

$$\max_{i\le n} ||\hat{\Pi}_i - \Pi_i|| = O_p(\xi_0(k)\triangle_{n,1}) \tag{23}$$
$$\max_{i\le n} ||\dot{\bar{\varphi}}_{l,i} - \bar{\varphi}_{l,i}|| = O_p(\xi_0(k)\triangle_{n,2}) \text{ for all } l. \tag{24}$$

Define $\hat{\mathcal{T}} = (\hat{\hat{\psi}}^{\mathbf{L},n})'\hat{\hat{\psi}}^{\mathbf{L},n}/n$ and $\dot{\mathcal{T}} = (\psi^{\mathbf{L},n})'\psi^{\mathbf{L},n}/n$. Our goal is to show that $\hat{\mathcal{T}}$ is nonsingular w.p.a.1. We first show that $\dot{\mathcal{T}}$ is nonsingular w.p.a.1 and this is closely related with the identification result of Theorem 4. Recall that (x_i, z_{1i}) and $\kappa(z_i,v_i)$ has no additive functional relationship for any $\kappa(z_i,v_i)$ satisfying $E[\kappa(Z_i,V_i)|Z_i] = 0$ and $E[\tilde{\varphi}^L_i\tilde{\varphi}^{L\prime}_i]$ is nonsingular by Assumption L1 (ii). Therefore, $\dot{\mathcal{T}}$ is nonsingular w.p.a.1 by Assumption L1 (ii) as $\zeta_0(\mathbf{L})^2\mathbf{L}/n \to 0$ by the same proof in Lemma A1 of Newey, Powell, and Vella (1999). The same conclusion holds even when instead we take $\dot{\mathcal{T}} = \sum_{i=1}^n C(z_i,v_i)\psi^{\mathbf{L}}_i\psi^{\mathbf{L}\prime}_i/n$ for some positive bounded function $C(z_i,v_i)$ by the same proof in Lemma A1 of Newey, Powell, and Vella (1999) and this helps to derive the consistency of the heteroskedasticity robust variance estimator later.

For ease of notation along the proof, we will assume some rate conditions are satisfied. Then we collect those rate conditions in Section A.2 and derive conditions under which all of them are satisfied.

Next note that

$$\begin{aligned}\left\|\hat{\tilde{\varphi}}^L_i - \tilde{\varphi}^L_i\right\| &\le \left\|\varphi^L(z_i,\hat{v}_i) - \varphi^L(z_i,v_i)\right\| + \left\|\hat{\bar{\varphi}}^L(z_i) - \bar{\varphi}^L(z_i)\right\| \\ &\le \left\|\varphi^L(z_i,\hat{v}_i) - \varphi^L(z_i,v_i)\right\| + \left\|\hat{\bar{\varphi}}^L(z_i) - \dot{\bar{\varphi}}^L(z_i)\right\| + \left\|\dot{\bar{\varphi}}^L(z_i) - \bar{\varphi}^L(z_i)\right\|.\end{aligned} \tag{25}$$

We find $\left\|\varphi^L(z_i,\hat{v}_i)-\varphi^L(z_i,v_i)\right\| \le C\zeta_1(L)||\hat{\Pi}_i-\Pi_i||$ applying a mean value expansion because $\varphi_l(z_i,v_i)$ is Lipschitz in Π_i for all l (Assumption C1 (vi)). Combined with (22), it implies that

$$\sum_{i=1}^n \left\|\varphi^L(z_i,\hat{v}_i)-\varphi^L(z_i,v_i)\right\|^2/n = O_p(\zeta_1(L)^2\triangle_{n,1}^2). \tag{26}$$

Next let $\hat{\omega}_l=(\varphi_l(z_1,\hat{v}_1)-\varphi_l(z_1,v_1),\ldots,\varphi_l(z_n,\hat{v}_n)-\varphi_l(z_n,v_n))'$. Then we can write for any $l=1,\ldots,L$,

$$\begin{aligned}\sum_{i=1}^n \left\|\hat{\dot{\varphi}}_l(z_i)-\dot{\varphi}_l(z_i)\right\|^2/n &= \mathrm{tr}\left\{\sum_{i=1}^n p^k(z_i)'(P'P)^-P'\hat{\omega}_l\hat{\omega}_l'P(P'P)^-p^k(z_i)\right\}/n \\ &= \mathrm{tr}\left\{(P'P)^-P'\hat{\omega}_l\hat{\omega}_l'P(P'P)^-\sum_{i=1}^n p^k(z_i)p^k(z_i)'\right\}/n \\ &= \mathrm{tr}\left\{(P'P)^-P'\hat{\omega}_l\hat{\omega}_l'P\right\}/n \\ &\le C\max_{i\le n}||\hat{\Pi}_i-\Pi_i||^2\mathrm{tr}\left\{(P'P)^-P'P\right\}/n \le C\xi_0(k)^2\triangle_{n,1}^2k/n\end{aligned} \tag{27}$$

where the first inequality is obtained by (23) and applying a mean value expansion to $\varphi_l(z_i,v_i)$ which is Lipschitz in Π_i for all l (Assumption C1 (vi)). From (22), (25), (26), and (27), we conclude

$$\sum_{i=1}^n ||\hat{\bar{\varphi}}^L(z_i)-\bar{\varphi}^L(z_i)||^2/n = O_p(L\xi_0(k)^2\triangle_{n,1}^2k/n)+O_p(L\triangle_{n,2}^2)=o_p(1) \tag{28}$$

and

$$\sum_{i=1}^n \left\|\hat{\tilde{\varphi}}_i^L-\tilde{\varphi}_i^L\right\|^2/n = O_p(\zeta_1(L)^2\triangle_{n,1}^2)+O_p(L\xi_0(k)^2\triangle_{n,1}^2k/n)+O_p(L\triangle_{n,2}^2)=o_p(1). \tag{29}$$

This also implies that by the triangle inequality and the Markov inequality,

$$\sum_{i=1}^n ||\hat{\tilde{\varphi}}_i^L||^2/n \le 2\sum_{i=1}^n ||\hat{\tilde{\varphi}}_i^L-\tilde{\varphi}_i^L||^2/n+2\sum_{i=1}^n ||\tilde{\varphi}_i^L||^2/n = o_p(1)+O_p(L). \tag{30}$$

Let $\triangle_n^\varphi=(\zeta_1(L)+L^{1/2}\xi(k)\sqrt{k/n}+L^{1/2})\triangle_n$. It also follows that

$$\sum_{i=1}^n \left\|\hat{\psi}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|^2/n \le \sum_{i=1}^n \left\|\hat{\tilde{\varphi}}_i^L-\tilde{\varphi}_i^L\right\|^2/n = O_p((\triangle_n^\varphi)^2)=o_p(1). \tag{31}$$

This also implies

$$\sum_{i=1}^n \left\|\hat{\psi}_i^{\mathbf{L}}\right\|^2/n = O_p(\mathbf{L})$$

because $\sum_{i=1}^n \left\|\hat{\psi}_i^{\mathbf{L}}\right\|^2/n \le 2\sum_{i=1}^n \left\|\hat{\psi}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|^2/n+2\sum_{i=1}^n \left\|\psi_i^{\mathbf{L}}\right\|^2/n = O_p(\mathbf{L})$.

Then applying (31) and applying the triangle inequality and Cauchy-Schwarz inequality and by Assumption L1 (iii) , we obtain

$$\begin{aligned}||\hat{\mathcal{T}}-\dot{\mathcal{T}}|| &\leq \sum_{i=1}^{n}\left\|\hat{\hat{\psi}}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|^2/n+2\sum_{i=1}^{n}\left\|\psi_i^{\mathbf{L}}\right\|\left\|\hat{\hat{\psi}}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|/n \\ &\leq O_p((\triangle_n^{\varphi})^2)+2\left(\sum_{i=1}^{n}\left\|\psi_i^{\mathbf{L}}\right\|^2/n\right)^{1/2}\left(\sum_{i=1}^{n}\left\|\hat{\hat{\psi}}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|^2/n\right)^{1/2} \\ &= O_p((\triangle_n^{\varphi})^2)+O_p(\mathbf{L}^{1/2}\triangle_n^{\varphi})=o_p(1).\end{aligned} \tag{32}$$

It follows that

$$\begin{aligned}||\hat{\mathcal{T}}-\mathcal{T}|| &\leq ||\hat{\mathcal{T}}-\dot{\mathcal{T}}||+||\dot{\mathcal{T}}-\mathcal{T}|| \\ &= O_p((\triangle_n^{\varphi})^2+\mathbf{L}^{1/2}\triangle_n^{\varphi}+\zeta_0(\mathbf{L})\sqrt{\mathbf{L}/n})\equiv O_p(\triangle_{\mathcal{T}})=o_p(1)\end{aligned} \tag{33}$$

where we obtain $||\dot{\mathcal{T}}-\mathcal{T}||=O_p(\zeta_0(\mathbf{L})\sqrt{\mathbf{L}/n})$ by the same proof in Lemma A1 of Newey, Powell, and Vella (1999).

Therefore we conclude $\hat{\mathcal{T}}$ is also nonsingular w.p.a.1. The same conclusion holds even when instead we take $\hat{\mathcal{T}}=\sum_{i=1}^{n}C(z_i,v_i)\hat{\hat{\psi}}_i^{\mathbf{L}}\hat{\hat{\psi}}_i^{\mathbf{L}\prime}/n$ and $\dot{\mathcal{T}}=\sum_{i=1}^{n}C(z_i,v_i)\psi_i^{\mathbf{L}}\psi_i^{\mathbf{L}\prime}/n$ for some positive bounded function $C(z_i,v_i)$ and this helps to derive the consistency of the heteroskedasticity robust variance estimator later.

Let $\eta_i=y_i-g_0(z_i,v_i)$ and let $\eta=(\eta_1,\ldots,\eta_n)'$. Let $(\mathbf{Z},\mathbf{V})=((Z_1,V_1),\ldots,(Z_n,V_n))$. Then we have $E[\eta_i|\mathbf{Z},\mathbf{V}]=0$ and by the independence assumption of the observations, we have $E[\eta_i\eta_j|\mathbf{Z},\mathbf{V}]=0$ for $i\neq j$. We also have $E[\eta_i^2|\mathbf{Z},\mathbf{V}]<\infty$. Then by (31) and the triangle inequality, we bound

$$\begin{aligned}E\left[||(\hat{\hat{\psi}}^{\mathbf{L},n}-\psi^{\mathbf{L},n})'\eta/n||^2|\mathbf{Z},\mathbf{V}\right] &\leq Cn^{-2}\sum_{i=1}^{n}E[\eta_i^2|\mathbf{Z},\mathbf{V}]\left\|\hat{\hat{\psi}}_i^{\mathbf{L}}-\psi_i^{\mathbf{L}}\right\|^2 \\ &\leq n^{-1}O_p(L(\triangle_n^{\varphi})^2)=o_p(n^{-1}).\end{aligned}$$

Then from the standard result (see Newey (1997) or Newey, Powell, and Vella (1999)) that the bound of a term in the conditional mean implies the bound of the term itself, we obtain $||(\hat{\hat{\psi}}^{\mathbf{L},n}-\psi^{\mathbf{L},n})'\eta/n||^2=o_p(n^{-1})$. Also note that $E[\left\|(\psi^{\mathbf{L},n})'\eta/n\right\|^2]=C\mathbf{L}/n$ (see proof of Lemma A1 in Newey, Powell, and Vella (1999)). Therefore, by the triangle inequality

$$\begin{aligned}||(\hat{\hat{\psi}}^{\mathbf{L},n})'\eta/n||^2 &\leq 2||(\hat{\hat{\psi}}^{\mathbf{L},n}-\psi^{\mathbf{L},n})'\eta/n||^2+2||(\psi^{\mathbf{L},n})'\eta/n||^2. \\ &= o_p(1)+O_p(\mathbf{L}/n)=O_p(\mathbf{L}/n).\end{aligned} \tag{34}$$

Define

$$\hat{g}_i = \hat{f}(x_i,z_{1i})+\hat{h}(z_i,\hat{v}_i),\hat{\hat{g}}_{\mathbf{L}i}=f_K(x_i,z_{1i})+\hat{h}_L(z_i,\hat{v}_i),\tilde{g}_{\mathbf{L}i}=f_K(x_i,z_{1i})+h_L(z_i,\hat{v}_i),$$

$\tilde{g}_{0i} = f_0(x_i, z_{1i}) + h_0(z_i, \hat{v}_i)$, and $g_{0i} = f_0(x_i, z_{1i}) + h_0(z_i, v_i)$ where $f_K(x_i, z_{1i}) = \sum_{l=1}^K \beta_l \phi_l(x_i, z_{1i})$, $\hat{h}(z_i, \hat{v}_i) = \hat{a}_L' \hat{\hat{\varphi}}(z_i, \hat{v}_i)$, $\hat{h}_L(z_i, \hat{v}_i) = a_L' \hat{\hat{\varphi}}(z_i, \hat{v}_i)$, and $h_L(z_i, \hat{v}_i) = a_L'(\varphi(z_i, \hat{v}_i) - \bar{\varphi}^L(z_i))$ and let $\hat{g}, \hat{\hat{g}}_{\mathbf{L}}$, $\tilde{g}_{\mathbf{L}}$, and $\tilde{g}_0$ stack the n observations of $\hat{g}_i, \hat{\hat{g}}_{\mathbf{L}i}$, $\tilde{g}_{\mathbf{L}i}$, and $\tilde{g}_{0i}$, respectively. Recall $\beta_{\mathbf{L}} = (\beta_1, \ldots, \beta_K, a_L')'$ and let this $\beta_{\mathbf{L}}$ satisfies Assumption L1 (iv). From the first order condition of the last step least squares we obtain

$$\begin{aligned} 0 &= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(y - \hat{g})/n \qquad (35) \\ &= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(\eta - (\hat{g} - \hat{\hat{g}}_{\mathbf{L}}) - (\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}}) - (\tilde{g}_{\mathbf{L}} - \tilde{g}_0))/n \\ &= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(\eta - \hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\beta} - \beta_{\mathbf{L}}) - (\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}}) - (\tilde{g}_{\mathbf{L}} - \tilde{g}_0) - (\tilde{g}_0 - g_0))/n. \end{aligned}$$

Note that by $\hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\hat{\psi}}^{\mathbf{L},n\prime}\hat{\hat{\psi}}^{\mathbf{L},n})^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}$ idempotent and by Assumption L1 (iv),

$$\begin{aligned} \|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)/n\| &\leq O_p(1)\{(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)'\hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\hat{\psi}}^{\mathbf{L},n\prime}\hat{\hat{\psi}}^{\mathbf{L},n})^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)/n\}^{1/2} \qquad (36) \\ &\leq O_p(1)\{(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)'(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)/n\}^{1/2} = O_p(\mathbf{L}^{-\gamma}). \end{aligned}$$

Similarly we obtain by $\hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\hat{\psi}}^{\mathbf{L},n\prime}\hat{\hat{\psi}}^{\mathbf{L},n})^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}$ idempotent, Assumption L1 (iv), and (28),

$$\begin{aligned} &\|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}})/n\| = O_p(1)\{(\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}})'(\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}})/n\}^{1/2} \qquad (37) \\ \leq\ & O_p(1)(\sum_{i=1}^n \|\hat{h}_L(z_i, \hat{v}_i) - \tilde{h}_L(z_i, \hat{v}_i)\|^2/n)^{1/2} \\ \leq\ & O_p(1)(\sum_{i=1}^n \|a_L\|^2 \|\hat{\bar{\varphi}}^L(z_i) - \bar{\varphi}^L(z_i)\|^2/n)^{1/2} = O_p(L\xi_0(k)\triangle_{n,1}\sqrt{k/n} + L\triangle_{n,2}). \end{aligned}$$

Similarly also by $\hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\hat{\psi}}^{\mathbf{L},n\prime}\hat{\hat{\psi}}^{\mathbf{L},n})^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}$ idempotent and (22) and applying the mean value expansion to $h_0(z_i, v_i)$, we have

$$\begin{aligned} &\|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\tilde{g}_0 - g_0)/n\| = O_p(1)(\sum_{i=1}^n \|h_0(z_i, \hat{v}_i) - h_0(z_i, v_i)\|^2/n)^{1/2} \qquad (38) \\ \leq\ & O_p(1)(\sum\nolimits_{i=1}^n \|\hat{\Pi}_i - \Pi_i\|^2/n)^{1/2} = O_p(\triangle_{n,1}) = o_p(1). \end{aligned}$$

Combining (34), (35), (36), (37), (38) and by $\hat{\mathcal{T}}$ is nonsingular w.p.a.1, we obtain

$$\begin{aligned} \|\hat{\beta} - \beta_{\mathbf{L}}\| &\leq \|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}\eta/n\| + \|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{\hat{g}}_{\mathbf{L}} - \tilde{g}_{\mathbf{L}})/n\| + \|\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\tilde{g}_{\mathbf{L}} - \tilde{g}_0)/n\| + o_p(1) \\ &= O_p(1)\{\sqrt{\mathbf{L}/n} + L\xi_0(k)\triangle_{n,1}\sqrt{k/n} + L\triangle_{n,2} + \mathbf{L}^{-\gamma}\} \equiv O_p(\triangle_{n,\beta}). \qquad (39) \end{aligned}$$

Define $g^*_{\mathbf{L}i} = f_K(x_i, z_{1i}) + h^*_L(z_i, v_i)$ where $h^*_L(z_i, v_i) = a_L'(\varphi^L(z_i, v_i) - \hat{\bar{\varphi}}^L(z_i))$. Then applying the triangle inequality, by (28), (39), the Markov inequality, Assumption L1 (iv), and $\hat{\mathcal{T}}$ is nonsingular w.p.a.1 (by Assumption L1 (ii) and (33)), we conclude

$$\begin{aligned}
& \sum_{i=1}^{n} \left(\hat{g}(z_i, v_i) - g_0(z_i, v_i)\right)^2 / n \qquad (40) \\
\leq\ & 3\sum_{i=1}^{n} \left(\hat{g}(z_i, v_i) - g^*_{\mathbf{L}i}\right)^2 / n + 3\sum_{i=1}^{n} \left(g^*_{\mathbf{L}i} - g_{\mathbf{L}i}\right)^2 / n + 3\sum_{i=1}^{n} \left(g_{\mathbf{L}i} - g_0(z_i, v_i)\right)^2 / n \\
\leq\ & O_p(1)||\hat{\beta} - \beta_{\mathbf{L}}||^2 \\
& + C_1 \sum_{i=1}^{n} ||a_L||^2 ||\hat{\bar{\varphi}}^L(z_i) - \bar{\varphi}^L(z_i)||^2/n + C_2 \sup_{\mathcal{W}} ||\beta'_{\mathbf{L}}\psi^{\mathbf{L}}(z,v) - g_0(z,v)||^2 \\
\leq\ & O_p(\triangle^2_{n,\beta}) + LO_p(L\xi_0(k)^2\triangle^2_{n,1}k/n + L\triangle^2_{n,2}) + O_p(\mathbf{L}^{-2\gamma}) = O_p(\triangle^2_{n,\beta}).
\end{aligned}$$

This also implies that by a similar proof to Theorem 1 of Newey (1997)

$$\max_{i \leq n} |\hat{g}_i - g_{0i}| = O_p(\zeta_0(\mathbf{L})\triangle_{n,\beta}). \qquad (41)$$

A.2 Proof of Theorem 5

Under Assumptions C1, all the assumptions in Assumption L1 are satisfied. For the consistency, we require the following rate conditions: R(i) $\mathbf{L}^{1/2}\triangle^{\varphi}_n \to 0$ from (32), R(ii) $\zeta_0(\mathbf{L})^2\mathbf{L}/n \to 0$ (such that $\dot{\mathcal{T}}$ is nonsingular w.p.a.1), and R(iii) $\xi_0(k)^2 k/n \to 0$ (such that $P'P/n$ is nonsingular w.p.a.1). The other rate conditions are dominated by these three. From the definition of $\triangle^{\varphi}_n = (\zeta_1(L) + L^{1/2}\xi_0(k)\sqrt{k/n} + L^{1/2})\triangle_n$, we have R(i) : $\mathbf{L}^{1/2}(\zeta_1(L) + L^{1/2}\xi_0(k)\sqrt{k/n} + L^{1/2})\triangle_n$.

For the polynomial approximations, we have $\zeta_\delta(L) \leq CL^{1+2\delta}$ and $\xi_0(k) \leq Ck$ and for the spline approximations, we have $\zeta_\delta(L) \leq CL^{0.5+\delta}$ and $\xi_0(k) \leq Ck^{0.5}$. Therefore for the polynomial approximations, the rate condition becomes (i) $\mathbf{L}^{1/2}(L^3 + L^{1/2}k^{3/2}/\sqrt{n} + L^{1/2})\triangle_n \to 0$, (ii) $\mathbf{L}^3/n \to 0$, and (iii) $k^3/n \to 0$ and for the spline approximations, it becomes R(i) $\mathbf{L}^{1/2}(L^{3/2} + L^{1/2}k/\sqrt{n} + L^{1/2})\triangle_n \to 0$, (ii) $\mathbf{L}^2/n \to 0$, and (iii) $k^2/n \to 0$. Also note that

$$\begin{aligned}
\triangle_{n,\beta} &\equiv \sqrt{\mathbf{L}/n} + L\xi_0(k)\triangle_{n,1}\sqrt{k/n} + L\triangle_{n,2} + \mathbf{L}^{-\gamma} \\
&= \sqrt{\mathbf{L}/n} + L\triangle_n + \mathbf{L}^{-\gamma}
\end{aligned}$$

since $\xi_0(k)\sqrt{k/n} = o(1)$. We take $\gamma = s/d$ because f_0 and h_0 belong to the Hölder class and we can apply the approximation theorems (e.g., see Timan (1963), Schumaker (1981), Newey (1997), and Chen (2007)).

Therefore, the conclusion of Theorem C1 follows from Lemma L1 applying the dominated convergence theorem by $\hat{g}_i$ and g_{0i} are bounded.

B Proof of asymptotic normality

Along the proof, we will obtain a series of convergence rate conditions. We collect them here. First define

$$
\begin{aligned}
\triangle_n^{\varphi} &= (\zeta_1(L) + L^{1/2}\xi_0(k)\sqrt{k/n} + L^{1/2})\triangle_n \\
\triangle_{n,\beta} &= \sqrt{\mathbf{L}/n} + L\triangle_n + \mathbf{L}^{-\gamma} \\
\triangle_{\mathcal{T}} &= (\triangle_n^{\varphi})^2 + \mathbf{L}^{1/2}\triangle_n^{\varphi} + \zeta_0(\mathbf{L})\sqrt{\mathbf{L}/n}, \triangle_{\mathcal{T}_1} = \xi_0(k)\sqrt{k/n} \\
\triangle_H &= \zeta_0(\mathbf{L})k^{1/2}/\sqrt{n} + k^{1/2}\triangle_n^{\varphi} + L^{-\gamma}\zeta_0(\mathbf{L})\sqrt{k} \\
\triangle_{d\varphi} &= \zeta_0(\mathbf{L})L^{1/2}\triangle_{n,2}, \triangle_g = \zeta_0(\mathbf{L})\triangle_{n,\beta} \\
\triangle_{\Sigma} &= \boldsymbol{\Delta}_{\mathcal{T}} + \zeta_0(\mathbf{L})^2\mathbf{L}/n, \triangle_{\hat{H}} = (\zeta_1(L)\triangle_{n,\beta} + \xi_0(k)\triangle_{n,1})\mathbf{L}^{1/2}\xi_0(k)
\end{aligned}
$$

and we need the following rate conditions for the $\sqrt{n}$-consistency and the consistency of the variance matrix estimator $\hat{\Omega}$:

$$
\begin{gathered}
\sqrt{n}\mathbf{L}^{-\gamma} \to 0, \sqrt{n}k^{1/2}L^{-\gamma} \to 0, \sqrt{n}k^{-\gamma_1} \to 0, \sqrt{n}k^{-\gamma_2} \to 0 \\
k^{1/2}(\triangle_{\mathcal{T}_1} + \triangle_H) + \mathbf{L}^{1/2}\triangle_{\mathcal{T}} \to 0, n^{-1}(\zeta_0(\mathbf{L})^2\mathbf{L} + \xi_0(k)^2k + \xi_0(k)^2kL^4) \to 0, \\
k^{1/2}(\triangle_{\mathcal{T}_1} + \triangle_H) + \mathbf{L}^{1/2}\triangle_{\mathcal{T}} + \triangle_{d\varphi} \to 0, \triangle_g \to 0, \triangle_{\Sigma} \to 0, \triangle_{\hat{H}} \to 0.
\end{gathered}
$$

Dropping the dominated ones and assuming $\sqrt{n}\mathbf{L}^{-\gamma}$, $\sqrt{n}k^{-\gamma_1}$, and $\sqrt{n}k^{-\gamma_2}$ are small enough, under the following all the rate conditions are satisfied:

$$
\frac{\zeta_0(\mathbf{L})k + \zeta_1(L)k^{3/2} + \zeta_0(\mathbf{L})\mathbf{L} + \mathbf{L}\zeta_1(L)\xi_0(k) + \mathbf{L}^{1/2}\zeta_1(L)L\xi_0(k)k^{1/2} + \mathbf{L}^{1/2}\xi_0(k)^2k^{1/2}}{\sqrt{n}} \to 0
$$

for the polynomial approximations it becomes $\frac{\mathbf{L}^2+\mathbf{L}L^3k+\mathbf{L}^{1/2}(L^4k^{3/2}+k^{5/2})}{\sqrt{n}} \to 0$ and for the spline approximations it becomes $\frac{\mathbf{L}^{3/2}+\mathbf{L}L^{3/2}k^{1/2}+\mathbf{L}^{1/2}(L^{5/2}k+k^{3/2})+L^{3/2}k^{3/2}}{\sqrt{n}} \to 0.$

Let $p_i^k = p^k(Z_i)$. We start with introducing additional notation:

$$\begin{aligned}
\Sigma &= E[\psi_i^{\mathbf{L}}\psi_i^{\mathbf{L}\prime}\mathrm{var}(Y_i|Z_i,V_i)], \mathcal{T} = E[\psi_i^{\mathbf{L}}\psi_i^{\mathbf{L}\prime}], \mathcal{T}_1 = E[p_i^k p_i^{k\prime}], \quad (42)\\
\Sigma_1 &= E[V_i^2 p_i^k p_i^{k\prime}], \Sigma_{2,l} = E[(\varphi_l(Z_i,V_i) - \bar{\varphi}_l(Z_i))^2 p_i^k p_i^{k\prime}],\\
H_{11} &= E[\frac{\partial h_{0i}}{\partial V_i}\psi_i^{\mathbf{L}} p_i^{k\prime}], \bar{H}_{11} = \sum_{i=1}^n \frac{\partial h_{0i}}{\partial V_i}\psi_i^{\mathbf{L}} p_i^{k\prime}/n\\
H_{12} &= E[E[\frac{\partial h_{0i}}{\partial V_i}|Z_i]\psi_i^{\mathbf{L}} p_i^{k\prime}], \bar{H}_{12} = \sum_{i=1}^n E[\frac{\partial h_{0i}}{\partial V_i}|Z_i]\psi_i^{\mathbf{L}} p_i^{k\prime}/n\\
H_{2,l} &= E[a_l\psi_i^{\mathbf{L}} p_i^{k\prime}], \bar{H}_{2,l} = \sum_{i=1}^n a_l\psi_i^{\mathbf{L}} p_i^{k\prime}/n, H_1 = H_{11} - H_{12}, \bar{H}_1 = \bar{H}_{11} - \bar{H}_{12}\\
\bar{\Omega} &= \mathcal{A}\mathcal{T}^{-1}[\Sigma + H_1\mathcal{T}_1^{-1}\Sigma_1\mathcal{T}_1^{-1}H_1' + \textstyle\sum_{l=1}^L H_{2,l}\mathcal{T}_1^{-1}\Sigma_{2,l}\mathcal{T}_1^{-1}H_{2,l}']\mathcal{T}^{-1}\mathcal{A}'.
\end{aligned}$$

We let $\mathcal{T}_1 = E[p_i^k p_i^{k\prime}] = I$ and $E[\tilde{\varphi}_i{}^L\tilde{\varphi}_i{}^{L\prime}] = I$ without loss of generality.

Then $\bar{\Omega} = \mathcal{A}\mathcal{T}^{-1}\left[\Sigma + H_1\Sigma_1 H_1' + \sum_{l=1}^L H_{2,l}\Sigma_{2,l}H_{2,l}'\right]\mathcal{T}^{-1}\mathcal{A}'$. Let Γ be a symmetric square root of $\bar{\Omega}$. Because $\mathcal{T}$ is nonsingular and $\mathrm{var}(Y_i|Z_i,V_i)$ is bounded away from zero, $\Sigma - CI$ is positive semidefinite for some positive constant C. It follows that

$$\begin{aligned}
||\Gamma\mathcal{A}\mathcal{T}^{-1}|| &= \{\mathrm{tr}(\Gamma\mathcal{A}\mathcal{T}^{-1}\mathcal{T}^{-1}\mathcal{A}'\mathcal{T}')\}^{1/2} \le C\{\mathrm{tr}(\Gamma\mathcal{A}\mathcal{T}^{-1}\Sigma\mathcal{T}^{-1}\mathcal{A}'\Gamma')\}^{1/2}\\
&\le \{\mathrm{tr}(C\Gamma\bar{\Omega}\Gamma')\}^{1/2} \le C.
\end{aligned}$$

Next we show $\bar{\Omega} \to \Omega$. Under Assumption R1, we have $\mathcal{A} = E[\nu^*(Z,V)\psi_i^{\mathbf{L}\prime}]$. Take $\nu_{\mathbf{L}}^*(Z,V) = \mathcal{A}\mathcal{T}^{-1}\psi_i^{\mathbf{L}}$. Then note $E[||\nu^*(Z,V) - \nu_{\mathbf{L}}^*(Z,V)||^2] \to 0$ because (i) $\nu_{\mathbf{L}}^*(Z,V) = E[\nu^*(Z,V)\psi_i^{\mathbf{L}\prime}]\mathcal{T}^{-1}\psi_i^{\mathbf{L}}$ is a mean-squared projection of $\nu^*(z_i,v_i)$ on $\psi_i^{\mathbf{L}}$; (ii) $\nu^*(z_i,v_i)$ is smooth and the second moment of $\nu^*(z_i,v_i)$ is bounded, so it is well-approximated in the mean-squared error as assumed in Assumption R1. Let $\nu_i^* = \nu^*(Z_i,V_i)$ and $\nu_{\mathbf{L}i}^* = \nu_{\mathbf{L}}^*(Z_i,V_i)$. It follows that

$$\begin{aligned}
E[\nu_{\mathbf{L}i}^*\mathrm{var}(Y_i|Z_i,V_i)\nu_{\mathbf{L}i}^{*\prime}] &= \mathcal{A}\mathcal{T}^{-1}E[\psi_i^{\mathbf{L}}\mathrm{var}(Y_i|Z_i,V_i)\psi_i^{\mathbf{L}\prime}]\mathcal{T}^{-1}\mathcal{A}'\\
&\to E[\nu_i^*\mathrm{var}(Y_i|Z_i,V_i)\nu_i^{*\prime}].
\end{aligned}$$

It concludes that $\mathcal{A}\mathcal{T}^{-1}\Sigma\mathcal{T}^{-1}\mathcal{A}'$ converges to $E[\nu_i^*\mathrm{var}(Y_i|Z_i,V_i)\nu_i^{*\prime}]$ (the first term in Ω) as $k, K, L \to \infty$. Let

$$b_{\mathbf{L}i} = E[\nu_{\mathbf{L}i}^*\left(\frac{\partial h_{0i}}{\partial V_i} - E[\frac{\partial h_{0i}}{\partial V_i}|Z_i]\right)p_i^{k\prime}]p_i^k$$

and $b_i = E\left[\nu_i^*\left(\frac{\partial h_{0i}}{\partial V_i} - E[\frac{\partial h_{0i}}{\partial V_i}|Z_i]\right)p_i^{k\prime}\right]p_i^k$. Then $E[||b_{\mathbf{L}i} - b_i||^2] \le CE[||\nu_{\mathbf{L}i}^* - \nu_i^*||^2] \to 0$ where the first inequality holds because the mean square error of a least squares projection cannot be larger than the MSE of the variable being projected. Also note that $E[||\rho_v(Z_i)-b_i||^2]\to 0$ as $k\to\infty$ because b_i is a least squares projection of $\nu_i^*\left(\frac{\partial h_{0i}}{\partial V_i}-E\left[\frac{\partial h_{0i}}{\partial V_i}|Z_i\right]\right)$ on p_i^k and it converges to the conditional mean as $k\to\infty$. Finally note that

$$
\begin{aligned}
& E[b_{\mathbf{L}i}\mathrm{var}(V_i|Z_i)b_{\mathbf{L}i}'] \\
= \; & \mathcal{A}\mathcal{T}^{-1}E\left[\psi_i^{\mathbf{L}}\left(\frac{\partial h_{0i}}{\partial V_i}-E\left[\frac{\partial h_{0i}}{\partial V_i}|Z_i\right]\right)p_i^{k\prime}\right]E[\mathrm{var}(V_i|Z_i)p_i^k p_i^{k\prime}] \\
& \times E\left[p_i^k\left(\frac{\partial h_{0i}}{\partial V_i}-E\left[\frac{\partial h_{0i}}{\partial V_i}|Z_i\right]\right)\psi_i^{\mathbf{L}\prime}\right]\mathcal{T}^{-1}\mathcal{A}' \\
= \; & \mathcal{A}\mathcal{T}^{-1}H_1\Sigma_1 H_1'\mathcal{T}^{-1}\mathcal{A}'
\end{aligned}
$$

and this conclude that $\mathcal{A}\mathcal{T}^{-1}H_1\Sigma_1H_1'\mathcal{T}^{-1}\mathcal{A}'$ converges to $E[\rho_v(Z)\mathrm{var}(X|Z)\rho_v(Z)']$ (the second term in Ω). Similarly we can show that for all l

$$\mathcal{A}\mathcal{T}^{-1}H_{2,l}\Sigma_{2,l}H_{2,l}'\mathcal{T}^{-1}\mathcal{A}'\to E[\rho_{\bar{\varphi}_l}(Z)\mathrm{var}(\varphi_l(Z,V)|Z)\rho_{\bar{\varphi}_l}(Z)'].$$

Therefore we conclude $\bar{\Omega}\to\Omega$ as $k,K,L\to\infty$. This also implies that $\Gamma\to\Omega^{-1/2}$ and Γ is bounded.

Next we derive the asymptotic normality of $\sqrt{n}(\hat{\theta}-\theta_0)$. After we establish the asymptotic normality, we will show the convergence of the each term in (20) to the corresponding terms in (42). We show some of them first, which will be useful to derive the asymptotic normality. Note $||\hat{\mathcal{T}}-\mathcal{T}||=O_p(\triangle_{\mathcal{T}})=o_p(1)$ and $||\hat{\mathcal{T}}_1-\mathcal{T}_1||=O_p(\triangle_{\mathcal{T}_1})=o_p(1)$. We also have $||\Gamma\mathcal{A}(\hat{\mathcal{T}}^{-1}-\mathcal{T}^{-1})||=o_p(1)$ and $||\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1/2}||^2=O_p(1)$ (see proof in Lemma A1 of Newey, Powell, and Vella (1999)). We next show $||\bar{H}_{11}-H_{11}||=o_p(1)$. Let $H_{11\mathbf{L}}=E[\sum_{l=1}^L a_l\frac{\partial\varphi_l(Z_i,V_i)}{\partial V_i}\psi_i^{\mathbf{L}}p_i^{k\prime}]$ and $\bar{H}_{11\mathbf{L}}=\sum_{i=1}^n\sum_{l=1}^L a_l\frac{\partial\varphi_l(Z_i,V_i)}{\partial V_i}\psi_i^{\mathbf{L}}p_i^{k\prime}/n$. Similarly define $H_{12\mathbf{L}}$ and $\bar{H}_{12\mathbf{L}}$ and let $H_{1\mathbf{L}}=H_{11\mathbf{L}}-H_{12\mathbf{L}}$. By Assumption N1 (i), Assumption L1 (iii), and the Cauchy-Schwarz inequality,

$$
\begin{aligned}
& ||H_1-H_{1\mathbf{L}}||^2 \\
\leq \; & CE[||\{(\frac{\partial h_{0i}}{\partial V_i}-E[\frac{\partial h_{0i}}{\partial V_i}|Z_i])-\sum_l a_l(\frac{\partial\varphi_l(Z_i,V_i)}{\partial V_i}-E[\frac{\partial\varphi_l(Z_i,V_i)}{\partial V_i}|Z_i])\}\psi_i^{\mathbf{L}}p_i^{k\prime}||^2] \\
\leq \; & CL^{-2\gamma}E[||\psi_i^{\mathbf{L}}||^2\sum_k p_{ki}^2]=O(L^{-2\gamma}\zeta_0(\mathbf{L})^2k).
\end{aligned}
$$

Next consider that by Assumption L1 (iii) and the Cauchy-Schwarz inequality,

$$
\begin{aligned}
E[\sqrt{n}||\bar{H}_{11\mathbf{L}} - H_{11\mathbf{L}}||] &\leq C(E[(\sum_{l=1}^{L} a_l \frac{\partial \varphi_l(Z_i, V_i)}{\partial V_i})^2 ||\psi_i^{\mathbf{L}}||^2 \sum_k p_{ki}^2])^{1/2} \\
&= C(E[(\frac{\partial h_{Li}}{\partial V_i})^2 ||\psi_i^{\mathbf{L}}||^2 \sum_k p_{ki}^2])^{1/2} \leq C\zeta_0(\mathbf{L})k^{1/2}
\end{aligned}
$$

where the first equality holds because $\frac{\partial h_{Li}}{\partial V_i} = \sum_{l=1}^{L} a_l \frac{\partial \tilde{\varphi}_l(Z_i,V_i)}{\partial V_i} = \sum_{l=1}^{L} a_l \frac{\partial \varphi_l(Z_i,V_i)}{\partial V_i}$ and the last result holds because $h_{Li} \in \mathcal{H}_n$ (i.e. $|h_{Li}|_1$ is bounded). Similarly by (31), the Cauchy-Schwarz inequality, and the Markov inequality, we obtain

$$
\begin{aligned}
\left\|\bar{H}_{11} - \bar{H}_{11\mathbf{L}}\right\| &\leq Cn^{-1} \sum\nolimits_{i=1}^{n} |\sum_{l=1}^{L} a_l \frac{\partial \varphi_l(Z_i, V_i)}{\partial V_i}| \cdot ||\hat{\hat{\psi}}_i^{\mathbf{L}} - \psi_i^{\mathbf{L}}|| \cdot ||p_i^k|| \\
&\leq C \left(\sum\nolimits_{i=1}^{n} C_i ||\hat{\hat{\psi}}_i^{\mathbf{L}} - \psi_i^{\mathbf{L}}||^2/n\right)^{1/2} \cdot \left(\sum\nolimits_{i=1}^{n} ||p_i^k||^2/n\right)^{1/2} \leq O_p(k^{1/2}\triangle_n^{\varphi}).
\end{aligned}
$$

Therefore, we have $||\bar{H}_{11} - H_{11}|| = O_p(\zeta_0(\mathbf{L})k^{1/2}/\sqrt{n} + k^{1/2}\triangle_n^{\varphi} + L^{-\gamma}\zeta_0(\mathbf{L})\sqrt{k}) \equiv O_p(\triangle_H) = o_p(1)$. Similarly we can show that $||\bar{H}_{12} - H_{12}|| = o_p(1)$ and $||\bar{H}_{2,l} - H_{2,l}|| = o_p(1)$ for all l.

Now we derive the asymptotic expansion to obtain the influence functions. Further define $\hat{g}_{\mathbf{L}i} = f_K(x_i, z_{1i}) + \tilde{h}_L(z_i, \hat{v}_i)$ where $\tilde{h}_L(z_i, \hat{v}_i) = a_L'(\varphi^L(z_i, \hat{v}_i) - E[\varphi^L(Z_i, \hat{V}_i)|z_i])$ and $g_{\mathbf{L}i} = f_K(x_i, z_{1i}) + h_L(z_i, v_i)$. From the first order condition, we obtain the expansion similar to (35). Recall $\beta_{\mathbf{L}} = (\beta_1, \ldots, \beta_K, a_L')'$ and let this $\beta_{\mathbf{L}}$ satisfy Assumption N1 (i).

$$
\begin{aligned}
0 &= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(y - \hat{g})/\sqrt{n} \qquad (43)\\
&= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(\eta - (\hat{g} - \hat{\hat{g}}_{\mathbf{L}}) - (\hat{\hat{g}}_{\mathbf{L}} - \hat{g}_{\mathbf{L}}) - (\hat{g}_{\mathbf{L}} - g_{\mathbf{L}}) - (g_{\mathbf{L}} - g_0))/\sqrt{n} \\
&= \hat{\hat{\psi}}^{\mathbf{L},n\prime}(\eta - \hat{\hat{\psi}}^{\mathbf{L},n}(\hat{\beta} - \beta_{\mathbf{L}}) - (\hat{\hat{g}}_{\mathbf{L}} - \hat{g}_{\mathbf{L}}) - (\hat{g}_{\mathbf{L}} - g_{\mathbf{L}}) - (g_{\mathbf{L}} - g_0))/\sqrt{n}.
\end{aligned}
$$

Similar to (36), we obtain

$$
||\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(g_{\mathbf{L}} - g_0)/\sqrt{n}|| = O_p(\sqrt{n}\mathbf{L}^{-\gamma}). \qquad (44)
$$

Also note that

$$
\begin{aligned}
\sqrt{n}||\Gamma(\alpha(g_{\mathbf{L}}) - \alpha(g_0))|| &= \sqrt{n}||\Gamma|| \cdot ||\alpha(g_{\mathbf{L}} - g_0)|| \leq C\sqrt{n}\,\|\Gamma\| \cdot |\psi^{\mathbf{L}\prime}(\cdot)\beta_{\mathbf{L}} - g_0(\cdot)|_{\delta} \qquad (45)\\
&= O_p(\sqrt{n}\mathbf{L}^{-\gamma}) = o_p(1)
\end{aligned}
$$

because $\alpha(\cdot)$ is a linear functional and by Assumption N1 (i).

From the linearity of $\alpha(\cdot)$, (43), (44), and (45) we have

$$\begin{aligned}\sqrt{n}\Gamma(\hat{\theta}-\theta_0) &= \sqrt{n}\Gamma(\alpha(\hat{g})-\alpha(g_0)) = \sqrt{n}\Gamma(\alpha(\hat{g})-\alpha(g_{\mathbf{L}})) + \sqrt{n}\Gamma(\alpha(g_{\mathbf{L}})-\alpha(g_0)) \quad (46)\\ &= \sqrt{n}\Gamma\mathcal{A}(\hat{\beta}-\beta_{\mathbf{L}}) + \sqrt{n}\Gamma\{a(g_{\mathbf{L}})-a(g_0)\}\\ &= \Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\eta-(\hat{\hat{g}}_{\mathbf{L}}-\hat{g}_{\mathbf{L}})-(\hat{g}_{\mathbf{L}}-g_{\mathbf{L}}))/\sqrt{n}+o_p(1).\end{aligned}$$

Now we derive the stochastic expansion of $\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{g}_{\mathbf{L}}-g_{\mathbf{L}})/\sqrt{n}$. Note that by a second order mean-value expansion of each $\tilde{h}_{Li}$ around v_i,

$$\begin{aligned}&\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}(\hat{g}_{\mathbf{L}i}-g_{\mathbf{L}i})/\sqrt{n} = \Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}(\tilde{h}_{Li}-h_{Li})/\sqrt{n}\\ =\ &\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}(\frac{dh_{Li}}{dv_i}-E[\frac{dh_{Li}}{dV_i}|Z_i])(\hat{\Pi}_i-\Pi_i)/\sqrt{n}+\hat{\varsigma}\\ =\ &\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\bar{H}_1\hat{\mathcal{T}}_1^{-1}\sum\nolimits_{i=1}^{n}p_i^k v_i/\sqrt{n}+\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\bar{H}_1\hat{\mathcal{T}}_1^{-1}\sum\nolimits_{i=1}^{n}p_i^k(\Pi_i-p_i^{k\prime}\lambda_k^1)/\sqrt{n}\\ &+\Gamma A\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}(\frac{dh_{Li}}{dv_i}-E[\frac{dh_{Li}}{dV_i}|Z_i])(p_i^{k\prime}\lambda_k^1-\Pi_i)/\sqrt{n}+\hat{\varsigma}.\end{aligned}$$

and the remainder term $||\hat{\varsigma}||\leq C\sqrt{n}||\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1/2}||\zeta_0(L)\sum_{i=1}^n C_i||\hat{\Pi}_i-\Pi_i||^2/n = O_p(\sqrt{n}\zeta_0(L)\triangle_{n,1}^2) = o_p(1)$. Then by the essentially same proofs ((A.18) to (A.23)) in Lemma A2 of Newey, Powell, and Vella (1999), under $\sqrt{n}k^{-s_1/d_z}\to 0$ and $k^{1/2}(\triangle_{\mathcal{T}_1}+\triangle_H)+\mathbf{L}^{1/2}\triangle_{\mathcal{T}}\to 0$ (so that we can replace $\hat{\mathcal{T}}_1$ with $\mathcal{T}_1$, $\bar{H}_1$ with H_1, and $\hat{\mathcal{T}}$ with $\mathcal{T}$ respectively), we obtain

$$\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{g}_{\mathbf{L}}-g_{\mathbf{L}})/\sqrt{n} = \Gamma\mathcal{A}\mathcal{T}^{-1}H_1\sum\nolimits_{i=1}^{n}p_i^k v_i/\sqrt{n}+o_p(1). \quad (47)$$

This derives the influence function that comes from estimating v_i in the first step.

Next we derive the stochastic expansion of $\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{\hat{g}}_{\mathbf{L}}-\hat{g}_{\mathbf{L}})/\sqrt{n}$:

$$\begin{aligned}&\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}(\hat{\hat{g}}_{\mathbf{L}i}-\hat{g}_{\mathbf{L}i})/\sqrt{n} = \Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}a_L'(\hat{\bar{\varphi}}^L(z_i)-E[\varphi^L(Z_i,\hat{V}_i)|z_i])/\sqrt{n}\\ =\ &\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\{\sum\nolimits_l\bar{H}_{2,l}\hat{\mathcal{T}}_1^{-1}\sum\nolimits_{i=1}^{n}p_i^k\tilde{\varphi}_{li}/\sqrt{n}+\sum\nolimits_l\bar{H}_{2,l}\hat{\mathcal{T}}_1^{-1}\sum\nolimits_{i=1}^{n}p_i^k(\bar{\varphi}_l(z_i)-p_i^{k\prime}\lambda_{l,k}^2)/\sqrt{n}\}\\ &+\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}\sum\nolimits_l a_l(p_i^{k\prime}\lambda_{l,k}^2-\bar{\varphi}_l(z_i))/\sqrt{n}+\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\sum\nolimits_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}\rho_i/\sqrt{n} \quad (48)\end{aligned}$$

where $\rho_i = p_i^{k\prime}\hat{\mathcal{T}}_1^{-1}\sum_{i=1}^n p_i^k\sum_l a_l\{(\varphi_l(z_i,\hat{v}_i)-\varphi_l(z_i,v_i))-(E[\varphi_l(Z_i,\hat{V}_i)|z_i]-\bar{\varphi}_l(z_i))\}$. We focus on the last term in (48). Note that $p_i^{k\prime}\hat{\mathcal{T}}_1^{-1}\sum_{i=1}^n p_i^k(\varphi_l(z_i,\hat{v}_i)-\varphi_l(z_i,v_i))$ is a projection of $\varphi_l(z_i,\hat{v}_i)-\varphi_l(z_i,v_i)$ on p_i^k and it converges to the conditional mean $E[\varphi_l(Z_i,\hat{V}_i)|z_i]-\bar{\varphi}_l(z_i)$. Note that $E[\rho_i|Z_1,\ldots,Z_n]=0$ and therefore $E[||\rho_i||^2|Z_1,\ldots,Z_n]\leq LO_p(\triangle_{n,2}^2)$ by a similar proof to (22). It follows that by Assumption L1 (iii) and the Cauchy-Schwarz inequality,

$$E[||\sum_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}\rho_i/\sqrt{n}|||Z_1,\ldots,Z_n] \leq (E[||\hat{\hat{\psi}}_i^{\mathbf{L}}||^2||\rho_i||^2|Z_1,\ldots,Z_n])^{1/2} \leq C\zeta_0(\mathbf{L})L^{1/2}\triangle_{n,2}.$$

This implies that $\sum_{i=1}^{n}\hat{\hat{\psi}}_i^{\mathbf{L}}\rho_i/\sqrt{n} = O_p(\zeta_0(\mathbf{L})L^{1/2}\triangle_{n,2}) \equiv O_p(\triangle_{d\varphi}) = o_p(1)$.

Then again by the essentially same proofs ((A.18) to (A.23)) in Lemma A2 of Newey, Powell, and Vella (1999), under $\sqrt{n}k^{-s_2/d_z} \to 0$, $\sqrt{n}k^{1/2}L^{-s/d} \to 0$, and $k^{1/2}(\triangle_{\mathcal{T}_1} + \triangle_H) + \mathbf{L}^{1/2}\triangle_{\mathcal{T}} + \triangle_{d\varphi} \to 0$ (so that we can replace $\hat{\mathcal{T}}_1$ with $\mathcal{T}_1$, $\bar{H}_{2,l}$ with $H_{2,l}$, and $\hat{\mathcal{T}}$ with $\mathcal{T}$ respectively and we can ignore the last term in (48)), we obtain

$$\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}(\hat{\tilde{g}}_{\mathbf{L}} - \hat{g}_{\mathbf{L}})/\sqrt{n} = \Gamma\mathcal{A}\mathcal{T}^{-1}\sum\nolimits_l H_{2,l}\sum\nolimits_{i=1}^{n} p_i^k\tilde{\varphi}_{li}/\sqrt{n} + o_p(1). \tag{49}$$

This derives the influence function that comes from estimating $E[\varphi_{li}|Z_i]$'s in the middle step.

We can also show that replacing $\hat{\hat{\psi}}_i^{\mathbf{L}}$ with $\psi_i^{\mathbf{L}}$ does not influence the stochastic expansion by (31). Therefore by (46), (47), and (49), we obtain the stochastic expansion,

$$\sqrt{n}\Gamma(\hat{\theta}-\theta_0) = \Gamma\mathcal{A}\mathcal{T}^{-1}(\psi^{\mathbf{L},n\prime}\eta - H_1\sum\nolimits_{i=1}^{n}p_i^k v_i/\sqrt{n} - \sum\nolimits_l H_{2,l}\sum\nolimits_{i=1}^{n}p_i^k\tilde{\varphi}_{li}/\sqrt{n}) + o_p(1).$$

To apply the Lindeberg-Feller theorem, we check the Lindeberg condition. For any vector q with $||q|| = 1$, let $W_{in} = q'\Gamma\mathcal{A}\mathcal{T}^{-1}(\psi_i^{\mathbf{L}}\eta_i - H_1p_i^kv_i - \sum_l H_{2,l}p_i^k\tilde{\varphi}_{li})/\sqrt{n}$. Note that W_{in} is i.i.d, given n and by construction, $E[W_{in}] = 0$ and $\mathrm{var}(W_{in}) = 1/n$. Also note that $||\Gamma\mathcal{A}\mathcal{T}^{-1}|| \leq C$, $||\Gamma\mathcal{A}\mathcal{T}^{-1}H_j|| \leq C||\Gamma\mathcal{A}\mathcal{T}^{-1}|| \leq C$ by $CI - H_jH_j'$ being positive semidefinite for $j = 1, (2,1), \ldots, (2,L)$. Also note that $(\sum_{l=1}^{L}\tilde{\varphi}_{li})^4 \leq L^2(\sum_{l=1}^{L}\tilde{\varphi}_{li}^2)^2 \leq L^3\sum_{l=1}^{L}\tilde{\varphi}_{li}^4$. It follows that for any $\varepsilon > 0$,

$$\begin{aligned} & nE[1(|W_{in}| > \varepsilon)W_{in}^2] = n\varepsilon^2E[1(|W_{in}| > \varepsilon)(W_{in}/\varepsilon)^2] \leq n\varepsilon^{-2}E[|W_{in}|^4] \\ \leq\ & Cn\varepsilon^{-2}\{E[||\psi_i^{\mathbf{L}}||^4E[\eta_i^4|Z_i,V_i]] + E[||p_i^k||^4E[V_i^4|Z_i]] + L^3\sum_l E[||p_i^k||^4E[\tilde{\varphi}_{li}^4|Z_i]]\}/n^2 \\ \leq\ & Cn^{-1}(\zeta_0(\mathbf{L})^2\mathbf{L} + \xi_0(k)^2k + \xi_0(k)^2kL^4) = o(1). \end{aligned}$$

Therefore, $\sqrt{n}\Gamma(\hat{\theta}-\theta_0) \to_d N(0,I)$ by the Lindeberg-Feller central limit theorem. We have shown that $\bar{\Omega} \to \Omega$ and Γ is bounded. We therefore also conclude $\sqrt{n}(\hat{\theta}-\theta_0) \to_d N(0,\Omega^{-1})$.

Now we show the convergence of the each term in (20) to the corresponding terms in (42). Let $\hat{\eta}_i = y_i - \hat{g}(z_i,\hat{v}_i)$. Note that $\hat{\eta}_i^* \equiv \hat{\eta}_i^2 - \eta_i^2 = -2\eta_i(\hat{g}_i - g_{0i}) + (\hat{g}_i - g_{0i})^2$ and that $\max_{i\leq n}|\hat{g}_i - g_{0i}| = O_p(\zeta_0(\mathbf{L})\triangle_{n,\beta}) = o_p(1)$ by (41). Let $\hat{D} = \Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\hat{\hat{\psi}}^{\mathbf{L},n\prime}\mathrm{diag}\{1+|\eta_i|,\ldots,1+|\eta_n|\}\hat{\hat{\psi}}^{\mathbf{L},n}\hat{\mathcal{T}}^{-1}\mathcal{A}'\Gamma'$ and note that $\hat{\hat{\psi}}^{\mathbf{L},n}$ and $\hat{\mathcal{T}}$ only depend on $(Z_1,V_1),\ldots,(Z_n,V_n)$ and thus $E[\hat{D}|(Z_1,V_1),\ldots,(Z_n,V_n)] \leq C\Gamma\mathcal{A}\hat{\mathcal{T}}^{-1}\mathcal{A}'\Gamma' = O_p(1)$. Therefore, $||\hat{D}|| = O_p(1)$ as well. Next let $\tilde{\Sigma} = \sum_{i=1}^{n} \hat{\hat{\psi}}_i^{\mathbf{L}} \hat{\hat{\psi}}_i^{\mathbf{L}\prime} \eta_i^2/n$. Then,

$$\begin{aligned} ||\Gamma \mathcal{A} \hat{\mathcal{T}}^{-1}(\hat{\Sigma} - \tilde{\Sigma})\hat{\mathcal{T}}^{-1}\mathcal{A}'\Gamma'|| &= ||\Gamma \mathcal{A} \hat{\mathcal{T}}^{-1} \hat{\hat{\psi}}^{\mathbf{L},n\prime} \text{diag}\{\hat{\eta}_1^*, \ldots, \hat{\eta}_n^*\} \hat{\hat{\psi}}^{\mathbf{L},n} \hat{\mathcal{T}}^{-1}\mathcal{A}'\Gamma'|| \\ &\leq C \text{tr}(\hat{D}) \max_{i \leq n} |\hat{g}_i - g_{0i}| = O_p(1) o_p(1). \end{aligned} \tag{50}$$

Then, by the essentially same proof in Lemma A2 of Newey, Powell, and Vella (1999), we obtain

$$\begin{aligned} ||\tilde{\Sigma} - \Sigma|| &= O_p(\Delta_{\mathcal{T}} + \zeta_0(\mathbf{L})^2 \mathbf{L}/n) \equiv O_p(\triangle_{\Sigma}) = o_p(1), \quad (51) \\ ||\Gamma \mathcal{A} \hat{\mathcal{T}}^{-1}(\hat{\Sigma} - \Sigma)\hat{\mathcal{T}}^{-1}\mathcal{A}'\Gamma'|| &= o_p(1), \\ ||\Gamma \mathcal{A}(\hat{\mathcal{T}}^{-1}\Sigma\hat{\mathcal{T}}^{-1} - \mathcal{T}^{-1}\Sigma\mathcal{T}^{-1})\mathcal{A}'\Gamma'|| &= o_p(1). \end{aligned}$$

Then, by (50), (51), and the triangle ineq., we conclude $||\Gamma \mathcal{A} \hat{\mathcal{T}}^{-1} \hat{\Sigma} \hat{\mathcal{T}}^{-1} \mathcal{A}'\Gamma' - \Gamma \mathcal{A} \mathcal{T}^{-1} \Sigma \mathcal{T}^{-1} \mathcal{A}'\Gamma'|| = o_p(1)$. It remains to show that for $j = 1, (2,1), \ldots, (2, L)$,

$$\Gamma \mathcal{A}(\hat{\mathcal{T}}^{-1} \hat{H}_j \hat{\mathcal{T}}_1^{-1} \hat{\Sigma}_j \hat{\mathcal{T}}_1^{-1} \hat{H}_j' \hat{\mathcal{T}}^{-1} - \mathcal{T}^{-1} H_j \Sigma_j H_j' \mathcal{T}^{-1})\mathcal{A}'\Gamma' = o_p(1). \tag{52}$$

As we have shown $||\hat{\Sigma} - \Sigma|| = o_p(1)$, similarly we can show $||\hat{\Sigma}_j - \Sigma_j|| = o_p(1)$, $j = 1, (2,1), \ldots, (2, L)$.

We focus on showing $||\hat{H}_j - \bar{H}_j|| = o_p(1)$ for $j = 1, (2,1), \ldots, (2, L)$. First note that

$$||\hat{H}_{11} - \bar{H}_{11}|| = ||\sum_{i=1}^{n} (\sum_{l=1}^{L} \hat{a}_l \frac{\partial \varphi_l(z_i, \hat{v}_i)}{\partial v_i} - a_l \frac{\partial \varphi_l(z_i, v_i)}{\partial v_i}) \hat{\hat{\psi}}_i^{\mathbf{L}} p^k(z_i)'/n||$$

By the Cauchy-Schwarz inequality, (30), and Assumption L1 (iii), we have $\sum_{i=1}^{n} ||\hat{\hat{\psi}}_i^{\mathbf{L}} p_i^{k\prime}||^2/n \leq \sum_{i=1}^{n} ||\hat{\hat{\psi}}_i^{\mathbf{L}}||^2 ||p_i^k||^2/n = O_p(\mathbf{L}\xi_0(k)^2)$. Also note that by the triangle inequality, the Cauchy-Schwarz inequality, and by Assumption C1 (vi) and (23), applying a mean value expansion to $\frac{\partial \varphi_l(z_i, v_i)}{\partial v_i}$ w.r.t v_i,

$$\begin{aligned} & \sum_{i=1}^{n} ||\sum_{l=1}^{L} (\hat{a}_l \frac{\partial \varphi_l(z_i, \hat{v}_i)}{\partial v_i} - a_l \frac{\partial \varphi_l(z_i, v_i)}{\partial v_i})||^2/n \\ \leq\ & 2\sum_{i=1}^{n} ||\sum_{l=1}^{L} (\hat{a}_l - a_l) \frac{\partial \varphi_l(z_i, v_i)}{\partial v_i}||^2/n + 2\sum_{i=1}^{n} ||\sum_{l=1}^{L} \hat{a}_l (\frac{\partial \varphi_l(z_i, \hat{v}_i)}{\partial v_i} - \frac{\partial \varphi_l(z_i, v_i)}{\partial v_i})||^2/n \\ \leq\ & C||\hat{a} - a_L||^2 \sum_{i=1}^{n} ||\frac{\partial \tilde{\varphi}^L(z_i, v_i)}{\partial v_i}||^2/n + C_1 \sum_{i=1}^{n} ||\sum_{l=1}^{L} \hat{a}_l \frac{\partial^2 \varphi_l(z_i, \tilde{v}_i)}{\partial v_i^2} (\hat{\Pi}_i - \Pi_i)||^2/n \\ \leq\ & C||\hat{a} - a_L||^2 \sum_{i=1}^{n} ||\frac{\partial \tilde{\varphi}^L(z_i, v_i)}{\partial v_i}||^2/n + C_1 \max_{1 \leq i \leq n} ||\hat{\Pi}_i - \Pi_i||^2 \cdot \sum_{i=1}^{n} ||\sum_{l=1}^{L} \hat{a}_l \frac{\partial^2 \varphi_l(z_i, \tilde{v}_i)}{\partial v_i^2}||^2/n \\ =\ & O_p(\zeta_1^2(L) \triangle_{n,\beta}^2 + \xi_0^2(k) \triangle_{n,1}^2) \end{aligned}$$

where $\tilde{v}_i$ lies between $\hat{v}_i$ and v_i, which may depend on l. We therefore conclude by the triangle inequality and the Cauchy-Schwarz inequality, $||\hat{H}_{11} - \bar{H}_{11}|| \leq O_p((\zeta_1(L)\triangle_{n,\beta} + \xi_0(k)\triangle_{n,1})\mathbf{L}^{1/2}\xi_0(k)) = O_p(\triangle_{\hat{H}}) = o_p(1)$. Similarly we can show that $||\hat{H}_{12} - \bar{H}_{12}|| = o_p(1)$ and $||\hat{H}_{2,l} - \bar{H}_{2,l}|| = o_p(1)$ $l = 1, \ldots, L$. We have shown that $||\bar{H}_j - H_j|| = o_p(1)$ for $j = 1, (2,1), \ldots, (2, L)$ previously. Therefore, $||\hat{H}_j - H_j|| = o_p(1)$ for $j = 1, (2,1), \ldots, (2, L)$. Then by the similar proof like (50) and (51), the conclusion (52) follows. From (52) finally note that by Γ is bounded, $||\hat{\Omega} - \bar{\Omega}|| \leq C||\Gamma\hat{\Omega}\Gamma' - \Gamma\bar{\Omega}\Gamma'|| = o_p(1)$.

References

BLUNDELL, R., AND J. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," *Advances in Economics and Econometrics:Theory and Applications, Eighth World Congress.*

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics, Volume 6.*

HALL, P., AND J. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904–2929.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1272.

HECKMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1261–1272.

MURPHY, K., AND R. TOPEL (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3, 370–379.

NEWEY, W. (1984): "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14, 201–206.

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

NEWEY, W., AND J. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

NEWEY, W., J. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equaitons Models," *Econometrica*, 67, 565–603.

SCHUMAKER, L. (1981): *Spline Fuctions:Basic Theory.*

SMITH, R., AND R. BLUNDELL (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54, 679–685.

TELSER, L. (1964): "Iterative Estimation of a Set of Linear Regression Equations," *Journal of the American Statistical Association*, 59, 845–862.

TIMAN, A. (1963): *Theory of Approximation of Functions of a Real Variable.*