AARHUS UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

# Alternative Asymptotics and the Partially Linear Model with Many Regressors

**Matias D. Cattaneo, Michael Jansson and Whitney K. Newey**

**CREATES Research Paper 2012-02**

Department of Economics and Business
Aarhus University
Bartholins Allé 10
DK-8000 Aarhus C
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

# Alternative Asymptotics and the Partially Linear Model with Many Regressors*

Matias D. Cattaneo† Michael Jansson‡ Whitney K. Newey§

July, 2010

Revised January, 2012

**Abstract**

Non-standard distributional approximations have received considerable attention in recent years. They often provide more accurate approximations in small samples, and theoretical improvements in some cases. This paper shows that the seemingly unrelated "many instruments asymptotics" and "small bandwidth asymptotics" share a common structure, where the object determining the limiting distribution is a V-statistic with a remainder that is an asymptotically normal degenerate U-statistic. This general structure can be used to derive new results. We employ it to obtain a new asymptotic distribution of a series estimator of the partially linear model when the number of terms in the series approximation possibly grows as fast as the sample size. This alternative asymptotic experiment implies a larger asymptotic variance than usual. When the disturbance is homoskedastic, this larger variance is consistently estimated by any of the usual homoskedastic-consistent estimators provided a "degrees-of-freedom correction" is used. Under heteroskedasticity of unknown form, however, none of the commonly used heteroskedasticity-robust standard-error estimators are consistent under the "many regressors asymptotics". We characterize the source of this failure, and we also propose a new standard-error estimator that is consistent under both heteroskedasticity and "many regressors asymptotics". A small simulation study shows that these new confidence intervals have reasonably good empirical size in finite samples.

**JEL classification**: C13, C31.

**Keywords**: partially linear model, many terms, adjusted variance.

*The authors thank comments from Alfonso Flores-Lagunes, Lutz Kilian, seminar participants at Bristol, Brown, Exeter, Indiana, Michigan, MSU, Princeton, Rutgers, UCLA, UCSD, UC-Irvine, USC, Warwick and Yale, and conference participants at the 2010 Joint Statistical Meetings and the 2010 LACEA Impact Evaluation Network Conference. The first author gratefully acknowledges financial support from the National Science Foundation (SES 1122994). The second author gratefully acknowledges financial support from the National Science Foundation (SES 1124174) and the research support of CREATES (funded by the Danish National Research Foundation). The third author gratefully acknowledges financial support from the National Science Foundation (SES 1132399).

†Department of Economics, University of Michigan.

‡Department of Economics, University of California at Berkeley, and CREATES.

§Corresponding Author. Department of Economics, MIT.

## 1 Introduction

Many instrument asymptotics, where the number of instruments grows as fast as the sample size, has proven useful for instrumental variables (IV) estimators. Kunitomo (1980) and Morimune (1983) derived asymptotic variances that are larger than the usual formulae when the number of instruments and sample size grow at the same rate, and Bekker (1994) and others provided consistent estimators of these larger variances. Hansen, Hausman, and Newey (2008) showed that using many instrument standard errors provides a theoretical improvement for a range of number of instruments and a practical improvement for estimating the returns to schooling. Thus, many instrument asymptotics and the associated standard errors have been demonstrated to be a useful alternative to the usual asymptotics for instrumental variables.

Instrumental variable estimators implicitly depend on a nonparametric series estimator. Many instrument asymptotics has the number of series terms growing so fast that the series estimator is not consistent. Analogous asymptotics for kernel-density weighted average derivative estimators has been considered by Cattaneo, Crump and Jansson (2010, 2011). They show that when the bandwidth shrinks faster than needed for consistency of the kernel estimator the variance of the estimator is larger than the usual formula. They also find that correcting the variance provides an improvement over standard asymptotics for a range of bandwidths.

The purpose of this paper is to show that these results share a common structure and that this structure can be used to derive new results. The common structure is that the object determining the limiting distribution is a V-statistic, which can be decomposed into a bias term, a sample average, and a "remainder" that is an asymptotically normal degenerate U-statistic. Asymptotic normality of the remainder distinguishes this setting from other ones involving V-statistics. Here the asymptotically normal remainder comes from the number of series terms going to infinity, or bandwidth shrinking to zero, while the behavior of a degenerate U-statistic is more complicated in other settings. When the number of terms grows as fast as the sample size (or the bandwidth shrinks to zero at an appropriate rate) the remainder has the same magnitude as the leading term, resulting in an asymptotic variance larger than just the variance of the leading term. The many instrument and small bandwidth results share this structure. In keeping with this common structure, we will henceforth refer to such results under the general heading of "alternative asymptotics".

Applying this common structure to a series estimator of the partially linear model leads to new results. These results allow the number of terms in the series approximation to grow as fast as the sample size. The asymptotic distribution of the estimator is derived and it is shown to have a larger asymptotic variance than the usual formula. When the disturbance is homoskedastic this larger variance is consistently estimated by using the usual homoskedasticity consistent estimator with the proper degrees of freedom correction.[1] This result provides a large sample justification for the use of a degrees of freedom correction without normality of disturbances. It is also found that the White (1980) variance estimator is inconsistent with many regressors, being too small when the disturbance is homoskedastic. We propose a new variance estimator that is heteroskedasticity consistent when the number of series terms grows as fast as the sample size. A small-scale simulation study provides evidence supporting our theoretical findings and, in particular, shows that the new heteroskedasticity consistent variance estimator performs well across a large range of number of series terms in small samples. These results suggest that the new standard errors should be useful for inference in the partially linear model with many regressors. These findings complement the recent work of Stock and Watson (2007), who showed that the conventional heteroskedasticity-robust variance matrix estimator is biased when employed in the context of a linear panel data model with fixed-effects and gave a consistent estimator.

The rest of the paper is organized as follows. Section 2 describes the common structure of many instrument and small bandwidth asymptotics, and also shows how the structure leads to new results for the partially linear model. Section 3 formalizes the new distributional approximation for the partially linear model, while Section 4 discusses the corresponding asymptotic variance estimation. Section 5 concludes. The supplemental appendix reports the results from a small Monte Carlo experiment, and contains the proofs of our theoretical results.

[1] Under the assumption of homoskedasticity, similar results are obtained in Calhoun (2011) for the F-test in a linear model with the number of regressors growing as fast as the sample size, and in Kolesar, Chetty, Friedman, Glaeser, and Imbens (2011) for a Gaussian IV model with many invalid instruments.

## 2 A Common Structure

To describe the common structure of many instrument and small bandwidth asymptotics, let $W_1, ..., W_n$ denote independent data observations. We consider an estimator $\hat{\mu}$ (of $\mu_0$, a generic parameter of interest) satisfying:

$$\sqrt{n}\,(\hat{\mu} - \mu_0) = \hat{\Gamma}^{-1} S_n, \qquad S_n = \sum_{1 \leq i,j \leq n} u_{ij}^n(W_i, W_j), \tag{1}$$

where $u_{ij}^n(w_i, w_j)$ is a function of a pair of observations that can depend on $i$, $j$, and $n$. We allow $u$ to depend on $n$ to account for number of terms or bandwidths that change with the sample size. Also, we allow $u$ to vary with $i$ and $j$ to account for dependence on variables that are being conditioned on in the asymptotics, and so treated as nonrandom.

We assume throughout this section that there exists a non-random matrix $\Gamma_n$ satisfying $\Gamma_n^{-1}\hat{\Gamma} \to_p I$ for $I$ the identity matrix of the appropriate dimension, and hence we focus on the V-statistic $S_n$. (All limits are taken as $n \to \infty$ unless explicitly stated otherwise.) This V-statistic has a well known (Hoeffding-type) decomposition that we describe here because it is an essential feature of the common structure. For notational implicitly we will drop the $W_i$ and $W_j$ arguments and set $u_{ij}^n = u_{ij}^n(W_i, W_j)$ and $\tilde{u}_{ij}^n = u_{ij}^n + u_{ji}^n - \mathbb{E}[u_{ij}^n + u_{ji}^n]$. Let $\|\cdot\|$ denote the Euclidean norm.

**Proposition 1.** If $\mathbb{E}[\|u_{ij}^n\|] < \infty$ for all $i, j, n$, then

$$S_n = \Psi_n + U_n + B_n, \tag{2}$$

where

$$\Psi_n = \sum_{1 \leq i \leq n} \psi_i^n(W_i), \quad \psi_i^n(W_i) = u_{ii}^n - \mathbb{E}[u_{ii}^n] + \sum_{1 \leq j \leq n, j \neq i} \mathbb{E}[\tilde{u}_{ij}^n | W_i],$$

$$U_n = \sum_{2 \leq i \leq n} D_i^n(W_i, ..., W_1), \quad D_i^n(W_i, ..., W_1) = \sum_{1 \leq j \leq n, j < i} \left(\tilde{u}_{ij}^n - \mathbb{E}[\tilde{u}_{ij}^n | W_i] - \mathbb{E}[\tilde{u}_{ij}^n | W_j]\right),$$

$B_n = \mathbb{E}[S_n]$, and $\mathbb{E}[\psi_i^n(W_i)] = 0$, $\mathbb{E}[D_i^n(W_i, ..., W_1) | W_{i-1}, ..., W_1] = 0$, $\mathbb{E}[\Psi_n U_n] = 0$.

This result shows that $S_n$ can be decomposed into a sum of independent terms $\Psi_n$, a U-statistic remainder $U_n$ that is a martingale difference sum and uncorrelated with $\Psi_n$, and a pure bias term $B_n$. This decomposition of a V-statistic is well known (e.g., van der Vaart (1998, Chapter 11)), and is included here for exposition. It is important in many of the proofs of asymptotic normality of semiparametric estimators, including Powell, Stock, and Stoker (1989), with the limiting distribution being determined by $\Psi_n$, and $U_n$ being treated as a remainder that is of smaller order when the bandwidth shrinks slowly enough.

An interesting feature of this decomposition in semiparametric settings is that $U_n$ is asymptotically normal at some rate when the number of series terms grow or the bandwidth shrinks to zero. In other settings the asymptotic behavior of $U_n$ is more complicated. It is a degenerate U-statistic, that in general converges to a weighted sum of chi-squares (e.g., van der Vaart (1998, Chapter 12)). Apparently what occurs in semiparametric settings as the number of instruments grows or the bandwidth shrinks is that the individual contributions $D_i^n(W_i, ..., W_1)$ to $U_n$ are small enough to satisfy a Lindeberg-Feller condition. Combined with the martingale property of $U_n$, this leads to asymptotic normality of $U_n$. This asymptotic normality property of $U_n$ has been shown for both series and kernel estimators, as further explained below.

Alternative asymptotics occurs when the number of series terms grows or the bandwidth shrinks fast enough that $\Psi_n$ and $U_n$ have the same magnitude in the limit. Because of uncorrelatedness of $\Psi_n$ and $U_n$ the asymptotic variance will be larger than the usual formula which is $\lim_{n\to\infty} \mathbb{V}[\Psi_n]$. As a consequence, consistent variance estimation under alternative asymptotics requires accounting for the presence of $U_n$. Accounting for the presence of $U_n$ should also yield improvements when numbers of series terms and bandwidths do not satisfy the knife-edge conditions of alternative asymptotics. For instance, if the number of series terms grows just slightly slower than the sample size then accounting for the presence of $U_n$ should still give a better large sample approximation. Hansen, Hausman, and Newey (2008) show such an improvement for many instrument asymptotics. It would be good to consider such improved approximations more generally, though it is beyond the scope of this paper to do so.

We show next that both many instrument asymptotics and small bandwidth asymptotics have the structure described above, and we also exploit this approach to derive new results in the case of a series estimator of the partially linear model, which we refer to as "many terms asymptotics". Details for the first two examples are given in the supplemental appendix.

The first example is concerned with the case of many instrument asymptotics. For simplicity we focus on the jackknife IV estimator JIVE2 of Angrist, Imbens, and Krueger (1999), but the idea applies to other IV estimators such as the limited information maximum likelihood estimator. See Chao, Swanson, Hausman, Newey, and Woutersen (2012) for more details.

**Example 1: "Many Instrument Asymptotics".** Consider a linear structural equation $y_i = x_i'\beta_0 + \varepsilon_i$ with $\mathbb{E}[\varepsilon_i] = 0$, $i = 1, ..., n$, where $x_i$ is a vector and $y_i$ and $\varepsilon_i$ are scalar dependent variable and disturbance respectively. Let $z_i$ be a $K \times 1$ vector of instrumental variables that we treat as constants. Let $Q = Z(Z'Z)^{-}Z'$ denote the projection matrix on the column space of $Z = [z_1, ..., z_n]'$, and $Q_{ij}$ the $(i, j)$-th element of $Q$. For JIVE2,

$$\sqrt{n}(\hat{\beta} - \beta_0) = \hat{\Gamma}^{-1} \sum_{1 \le i,j \le n, i \ne j} Q_{ij} x_i \varepsilon_j / \sqrt{n}, \qquad \hat{\Gamma} = \sum_{1 \le i,j \le n, i \ne j} Q_{ij} x_i x_j' / n,$$

which is a special case of equation (1). Proposition 1, and symmetry of $Q$, implies that equation (2) is satisfied. In this case, $B_n = 0$ and

$$\begin{aligned} \Psi_n &= \sum_{1 \le i \le n} \Upsilon_i (1 - Q_{ii}) \varepsilon_i / \sqrt{n} + \sum_{1 \le i \le n} \left( \Upsilon_i - \sum_{1 \le j \le n} Q_{ij} \Upsilon_j \right) \varepsilon_i / \sqrt{n} \\ &= \sum_{1 \le i \le n} \Upsilon_i (1 - Q_{ii}) \varepsilon_i / \sqrt{n} + o_p(1), \end{aligned}$$

where $\Upsilon_i = \mathbb{E}[x_i]$ and the second equality will generally follow because $\Upsilon_i - \sum_{j=1}^n Q_{ij} \Upsilon_j$ is a regression residual that converges to zero as $K$ grows. Under standard asymptotics $K/n \to 0$, and hence $Q_{ii} \to 0$, so $\lim_{n \to \infty} \mathbb{V}[\Psi_n]$ coincides with the usual asymptotic IV variance. Finally, the degenerate U-statistic term is

$$U_n = \sum_{1 \le j < i \le n} Q_{ij} \left( v_i \varepsilon_j + v_j \varepsilon_i \right) / \sqrt{n}, \qquad v_i = x_i - \Upsilon_i,$$

which will also be asymptotically normal as $K \to \infty$ by the martingale central limit theorem under appropriate conditions, as shown by Chao, Swanson, Hausman, Newey, and Woutersen (2012). Alternative asymptotics occurs when $K$ grows as fast as $n$, resulting in $\Psi_n$ and $U_n$ having the same magnitude in the limit. ■

The next example shows that small bandwidth asymptotics for kernel estimators also has the structure outlined above. In this case we focus on the integrated squared density to keep the exposition simple and because it shares the common structure with the density-weighted average derivative estimator of Powell, Stock, and Stoker (1989) treated in Cattaneo, Crump, and Jansson (2011). This idea applies more generally to estimands defined as density-weigthed averages and ratios thereof (see, e.g., Newey, Hsieh, and Robins (2004, Section 2) and references therein).

**Example 2: "Small Bandwidth Asymptotics".** Consider the parameter $\mu_0 = \int f_0(w)^2 \mathrm{d}w = \mathbb{E}[f_0(W_i)]$ with $W_i$ a continuous random variable with p.d.f. $f_0$. A leave-one-out estimator is

$$\hat{\mu} = \sum_{1 \leq i,j \leq n, i \neq j} K_h(W_i - W_j)/n(n-1),$$

where $K(u)$ is a symmetric kernel and $K_h(u) = h^{-d}K(u/h)$. This estimator has the V-statistic form of equation (1) with $\hat{\Gamma} = 1$. Proposition 1, and symmetry of $K(u)$, implies that equation (2) is satisfied. In this case, under appropriate smoothness and kernel assumptions, $f_h(W_i) = \int K(u) f(W_i + hu) \mathrm{d}u$ will converge to $f_0(W_i)$ in mean square as $h \to 0$, so that $B_n = \sqrt{n}(\mathbb{E}[f_h(W_i)] - \mu_0)$ will be negligible and

$$\Psi_n = \sum_{1 \leq i \leq n} 2(f_h(W_i) - \mathbb{E}[f_h(W_i)])/\sqrt{n} = \sum_{1 \leq i \leq n} 2(f_0(W_i) - \mu_0)/\sqrt{n} + o_p(1).$$

This gives the well-known influence function $2(f_0(W_i) - \mu_0)$ for estimators of $\mu_0$, implying that $\lim_{n \to \infty} \mathbb{V}[\Psi_n]$ does correspond to the usual asymptotic variance for estimators of $\mu_0$. Finally, the degenerate U-statistic term is

$$U_n = 2 \sum_{1 \leq i < j \leq n} \{K_h(W_i - W_j) - f_h(W_i) - f_h(W_j) + \mathbb{E}[f_h(W_i)]\}/\sqrt{n}(n-1),$$

which will also be asymptotically normal by the martingale central limit theorem, under appropriate conditions, with convergence rate $nh^{d/2}$. In this example alternative asymptotics occurs when $h^d$ shrinks as fast as $1/n$, resulting in $\Psi_n$ and $U_n$ having the same magnitude in the limit. $\blacksquare$

The previous two examples show how several estimators share the common structure outlined above. We now show how this structure can be applied to derive new results. We study series estimation in the context of the partially linear model, an important and widely used model in empirical work (see, e.g., Imbens and Wooldridge (2009)). Our new results will shed light on the asymptotic behavior of this estimator, and the associated inference procedures, when the number of terms are allowed to grow as fast as the sample size.

**Example 3: "Many Terms Asymptotics".** Let $(y_i, x_i', z_i')'$, $i = 1, \dots, n$, be a random sample of the random vector $(y, x', z')'$, where $y \in \mathbb{R}$ is a dependent variable, and $x \in \mathbb{R}^{d_x \times 1}$ and $z \in \mathbb{R}^{d_z \times 1}$ are explanatory variables. The partially linear model is given by

$$y_i = x_i'\beta + g(z_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i|x_i, z_i] = 0, \qquad \sigma^2_\varepsilon(x_i, z_i) = \mathbb{E}[\varepsilon_i^2|x_i, z_i],$$

where $v_i = x_i - h(z_i)$ with $h(z_i) = \mathbb{E}[x_i|z_i]$. A series estimator of $\beta$ is obtained by regressing $y_i$ on $x_i$ and approximating functions of $z_i$. To describe the estimator, let $p^K(z) = (p_{1K}(z), \dots, p_{KK}(z))'$ be a vector of approximating functions, such as polynomials or splines, where $K$ denotes the number of terms in the regression. Here the unknown function $g(z)$ will be approximated by a linear combination of $p_{kK}(z)$. Therefore, letting $Y = [y_1, \cdots, y_n]' \in \mathbb{R}^{n\times 1}$, $X = [x_1, \cdots, x_n]' \in \mathbb{R}^{n\times d_x}$ and $P = [p^K(z_1), \dots, p^K(z_n)]'$, a series estimator of $\beta$ is given by

$$\hat{\beta} = (X'MX)^{-1}X'MY, \qquad M = I - Q, \qquad Q = P(P'P)^{-}P',$$

where $A^-$ denotes a generalized inverse of a matrix $A$ (satisfying $AA^-A = A$) and $X'MX$ will be non-singular with probability approaching one under appropriate conditions. Donald and Newey (1994) gave conditions for the asymptotic normality of this estimator using standard asymptotics, as discussed in more detail below. (See also, among others, Robinson (1988) and Linton (1995) for standard asymptotics results when using kernel estimators.)

Conditional on $Z = [z_1, ..., z_n]'$ this estimator has the structure outlined earlier, and for this reason we will condition on $Z$ throughout the following discussion so that expectations are implicitly conditional on $Z$. (Alternatively, we could assume that the regressors entering the partially linear model are non-random.) To explain how $\hat{\beta}$ fits within the common structure, let $h_i = h(z_i)$, $g_i = g(z_i)$, $G = [g_1, \cdots, g_n]'$, $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_n]'$, $H = [h_1, \cdots, h_n]'$, and $V = [v_1, \cdots, v_n]'$. By $Y = X\beta + G + \varepsilon$ we have

$$\sqrt{n}(\hat{\beta} - \beta) = (X'MX/n)^{-1}X'M(\varepsilon + G)/\sqrt{n} = \hat{\Gamma}^{-1}S_n \tag{3}$$

with

$$\hat{\Gamma} = \sum_{1 \le i,j \le n} M_{ij}x_i x_i'/n, \qquad S_n = \sum_{1 \le i,j \le n} x_i M_{ij}(g_j + \varepsilon_j)/\sqrt{n},$$

and where $M_{ij}$ represents the $(i, j)$-th element of the matrix $M$. This estimator has the V-statistic form of equation (1) with $W_i = (y_i, x_i')'$ and $\mu_0 = \beta$, $\hat{\Gamma} = X'MX/n$, and $u_{ij}^n(W_i, W_j) = x_i M_{ij}(g_j + \varepsilon_j)/\sqrt{n}$. By $\mathbb{E}[\varepsilon_i|x_i] = 0$ we have $\mathbb{E}[x_i\varepsilon_i] = 0$. Therefore, letting $u_{ij}^n = u_{ij}^n(W_i, W_j)$ as we have done previously, we have $\mathbb{E}[u_{ij}^n] = h_i M_{ij} g_j/\sqrt{n}$, $u_{ij}^n - \mathbb{E}[u_{ij}^n] = M_{ij}\left(v_i g_j + x_i\varepsilon_j\right)/\sqrt{n}$, $\tilde{u}_{ij} = M_{ij}\left(v_j g_i + v_i g_j + x_j\varepsilon_i + x_i\varepsilon_j\right)/\sqrt{n}$, and $\mathbb{E}[\tilde{u}_{ij}|W_i] = M_{ij}\left(v_i g_j + h_j\varepsilon_i\right)/\sqrt{n}$. As before, we obtain from Proposition 1 that equation (2) is satisfied. In this case, the main bias term $B_n = H'MG/\sqrt{n}$ will also be negligible under appropriate conditions, as shown in the next section. Moreover, by $M_{ii} = 1 - Q_{ii}$,

$$\begin{aligned}\Psi_n &= \sum_{1 \le i \le n} M_{ii} v_i \varepsilon_i/\sqrt{n} + \left(V'MG + H'M\varepsilon\right)/\sqrt{n} \\ &= \sum_{1 \le i \le n} M_{ii} v_i \varepsilon_i/\sqrt{n} + o_p(1),\end{aligned}$$

where the second equality will follow because the last term has mean zero and converges to zero in mean square as $K$ grows, as further discussed below. Under standard asymptotics $Q_{ii}$ will go to zero and hence the variance of $\Psi_n$ corresponds to the usual asymptotic variance.

Finally, noting that for $j < i$, $M_{ij} = -Q_{ij}$, we find that the degenerate U-statistic term is

$$U_n = -\sum_{1\leq i<j\leq n} Q_{ij}\left(v_i\varepsilon_j + v_j\varepsilon_i\right)/\sqrt{n}.$$

Remarkably this term is essentially the same as the degenerate U-statistic term for JIVE2 that was shown above. Consequently, the central limit theorem of Chao, Swanson, Hausman, Newey, and Woutersen (2012) is applicable to this problem. We will employ it to show that $U_n$ is asymptotically normal as $K \to \infty$, even when $K/n$ does not converge to zero. ■

This last example provides a new approach to studying the asymptotic distribution of semi-linear regression under many regressor asymptotics. This alternative asymptotic approximation is useful, for instance, when the number of covariates entering the nonparametric part is large relative to the sample size, as it is usually the case in observational studies and related empirical problems.

# 3 Many Regressor Asymptotics

In this section we make precise the discussion given in Example 3. In the next section we will present a consistent variance estimator and study the asymptotic properties of well-known variance estimators under different assumptions. To avoid possible confusion, in the rest of the paper we make explicit the conditioning set for all expectations.

The estimator $\hat{\beta}$ described in Example 3 may be intuitively interpreted as a two-step semiparametric estimator with tuning parameter $K$, because the unknown (regression) functions $g(\cdot)$ and $h(\cdot)$ are nonparametrically estimated in a preliminary step by the series estimator. Donald and Newey (1994) gave conditions for asymptotic normality of this estimator when $K/n \to 0$. Here we allow $K/n$ to be bounded away from zero under easy-to-interpret high-level assumptions.

It is important to control the bias from approximating unknown functions by a linear combination of $p^K(\cdot)$. The following first condition does so.

**Assumption 1.**

(a) For some $\alpha_h > 0$, there is $\eta_h$ such that $\sum_{i=1}^{n} \mathbb{E}[\|h(z_i) - \eta_h p^K(z_i)\|^2]/n = O(K^{-2\alpha_h})$.

(b) For some $\alpha_g > 0$, there is $\eta_g$ such that $\sum_{i=1}^{n} \mathbb{E}[(g(z_i) - p^K(z_i)'\eta_g)^2]/n = O(K^{-2\alpha_g})$.

These conditions are implied by conventional assumptions from approximation theory. When the support of $z_i$ is compact commonly used basis of approximation, such as polynomials or splines, will satisfy this assumption with $\alpha_h = s_h/d_z$ and $\alpha_g = s_g/d_z$, where $s_h$ and $s_g$ denotes the number of continuous derivatives of $h$ and $g$, respectively.

We also assume that certain conditional moments are bounded.

**Assumption 2.** There is $C < \infty$ such that $\mathbb{E}[\|v_i\|^4|z_i] \leq C$ and $\mathbb{E}[\varepsilon_i^4|z_i] \leq C$.

From equation (3) and the discussion in the previous section we see that the asymptotic distribution of $\hat{\beta}$ will be determined by the behavior of $\hat{\Gamma} = X'MX/n$ and $S_n = X'M(\varepsilon + G)/\sqrt{n}$. We consider the properties of each of these objects in turn.

For $\hat{\Gamma}$, it is shown in the supplementary appendix that, under Assumptions 1(a) and 2,

$$\hat{\Gamma} = \Gamma_n + o_p(K/n), \qquad \Gamma_n = \frac{1}{n} \sum_{1 \leq i \leq n} M_{ii} \mathbb{E}[v_i v_i'|z_i],$$

which characterizes the stochastic behavior of the matrix $\hat{\Gamma}$ without requiring that $K/n \to 0$. It follows from this result that as long as $K/n$ is bounded $\hat{\Gamma}$ will be close to $\Gamma_n$ in probability. In the homoskedastic $v_i$ case where $\mathbb{E}[v_i v_i'|z_i] = \mathbb{E}[v_i v_i']$ and where $\text{rank}(Q) = K$ we see that $\hat{\Gamma}$ is close to $\Gamma_n = \mathbb{E}[v_i v_i']\text{trace}(M)/n = (1 - K/n)\,\mathbb{E}[v_i v_i']$. More generally, with heteroskedasticity, $\hat{\Gamma}$ will be close to the weighted average $\Gamma_n$.

This result includes standard asymptotics as a special case when $K/n \to 0$, where the law of large numbers, iterated expectations, and $\sum_{i=1}^{n} Q_{ii}/n = \text{trace}(Q)/n \leq K/n \to 0$ imply

$$\Gamma_n = \sum_{i=1}^{n} \mathbb{E}[v_i v_i'|z_i]/n - \sum_{i=1}^{n} Q_{ii} \mathbb{E}[v_i v_i'|z_i]/n + o_p(1) = \mathbb{E}[v_i v_i'] + o_p(1).$$

Next we study $S_n$. Under Assumptions 1–2, it is shown in the supplementary appendix that

$$\begin{aligned} X'M(\varepsilon + G)/\sqrt{n} &= \Psi_n + U_n + B_n \\ &= V'M\varepsilon/\sqrt{n} + O_p(K^{-\alpha_h} + K^{-\alpha_g}) + O_p(\sqrt{n}K^{-(\alpha_h + \alpha_g)}). \end{aligned}$$

This result places no restriction on the growth of $K$. Interestingly, the bias term $B_n$ involves approximation of both unknown functions $h$ and $g$, implying an implicit trade-off between smoothness conditions for $h$ and $g$. The bias condition $\sqrt{n}K^{-(\alpha_h+\alpha_g)} \to 0$ only requires that $\alpha_h + \alpha_g$ be large enough, but not necessarily $\alpha_h$ and $\alpha_g$ separately. It follows that if $\sqrt{n}K^{-(\alpha_h+\alpha_g)} \to 0$ (implying $K \to \infty$) we have

$$S_n = X'M(\varepsilon + G)/\sqrt{n} = \frac{1}{\sqrt{n}} \sum_{1\leq i,j\leq n} M_{ij} v_i \varepsilon_j + o_p(1),$$

as discussed in Example 3 above. Furthermore, a simple variance calculation yields

$$\Sigma_n := \mathbb{V}[V'M\varepsilon/\sqrt{n}|Z] = \frac{1}{n} \sum_{1\leq i\leq n} M_{ii}^2 \mathbb{E}[v_i v_i' \varepsilon_i^2 | z_i] + \frac{1}{n} \sum_{1\leq i,j\leq n, i\neq j} Q_{ij}^2 \mathbb{E}[v_i v_i' | z_i] \mathbb{E}[\varepsilon_j^2 | z_j].$$

Here the first term following the equality corresponds to the usual asymptotic approximation, while the second term adds an additional term that accounts for large $K$.

It is interesting to consider what happens in some special cases. Under homoskedasticity of $\varepsilon_i$, i.e. where $\mathbb{E}[\varepsilon_i^2|x_i, z_i] = \sigma_\varepsilon^2$,

$$\Sigma_n = \frac{\sigma_\varepsilon^2}{n} \sum_{1\leq i,j\leq n} M_{ij}^2 \mathbb{E}[v_i v_i' | z_i] = \frac{\sigma_\varepsilon^2}{n} \sum_{1\leq i\leq n} M_{ii} \mathbb{E}[v_i v_i' | z_i] = \sigma_\varepsilon^2 \Gamma_n,$$

because $\sum_{j=1}^n M_{ij}^2 = M_{ii}$. If in addition $\mathbb{E}[v_i v_i'|z_i] = \mathbb{E}[v_i v_i']$ and $\operatorname{rank}(Q) = K$ then $\Sigma_n = \sigma_\varepsilon^2 (1 - K/n) \mathbb{E}[v_i v_i']$. Also, if $K/n \to 0$ then by $\sum_{i,j} Q_{ij}^2/n = \sum_i Q_{ii}/n \leq K/n \to 0$ we have

$$\Sigma_n = \frac{1}{n} \sum_{1\leq i\leq n} M_{ii}^2 \mathbb{E}[v_i v_i' \varepsilon_i^2 | z_i] + o_p(1) = \mathbb{E}[v_i v_i' \varepsilon_i^2] + o_p(1),$$

which corresponds to the standard asymptotics limiting variance.

Using the previous derivations, together with an appropriate central limit theorem for quadratic forms, it can be established that $V'M\varepsilon/\sqrt{n}$ is asymptotically normal with $\Sigma_n$ as its asymptotic variance matrix. The following theorem collects and formalizes these results.

**Theorem 1.** Suppose Assumptions 1 and 2 are satisfied, and there is $0 < C < 1$ such that $Q_{ii} < C$,

$\lambda_{\min}(\Sigma_n) \geq 1/C$. If $\sqrt{n}K^{-\alpha_x-\alpha_g} \to 0$ and $K/n \to \alpha \in [0,1)$, then

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0, I_{d_x}), \qquad \Omega_n = \Gamma_n^{-1}\Sigma_n\Gamma_n^{-1}.$$

If, in addition, $\mathbb{E}\left[\varepsilon_i^2|x_i,z_i\right] = \sigma_\varepsilon^2 > 0$ then $\Omega_n = \sigma_\varepsilon^2\Gamma_n^{-1}$.

Theorem 1 shows that $\hat{\beta}$ is asymptotically normal when $K/n$ need not converge to zero. This asymptotic distributional result does not rely on asymptotic linearity, nor on the actual convergence of the matrices $\Gamma_n$ and $\Sigma_n$, and leads to a new (larger) asymptotic variance that captures terms that are assumed away by the classical result. The asymptotic distribution result of Donald and Newey (1994) is obtained as a special case where $K/n \to 0$. More generally, when $K/n$ does not converge to zero, the asymptotic variance will be larger than the usual formula because it accounts for the contribution of "remainder" $U_n$ in Proposition 1. For instance, when both $\varepsilon_i$ and $v_i$ are homoskedastic and $P'P$ is nonsingular, the asymptotic variance is

$$\Gamma_n^{-1}\Sigma_n\Gamma_n^{-1} = \sigma_\varepsilon^2\Gamma_n^{-1} = \frac{n}{n-K}\sigma_\varepsilon^2(\mathbb{E}[v_iv_i'])^{-1},$$

which is larger than the usual asymptotic variance $\sigma_\varepsilon^2\left(\mathbb{E}[v_iv_i']\right)^{-1}$ by the degrees of freedom correction $n/(n-K)$.

## 4 Asymptotic Variance Estimation

Consistent asymptotic variance estimation is useful for large sample inference. It is well known that if $\sqrt{n}\Omega_n^{-1/2}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,I)$, $\Omega_n$ is bounded with smallest eigenvalue bounded away from zero, and $\hat{\Omega}-\Omega_n \xrightarrow{p} 0$, then for any matrix $B$ with full row rank,

$$(B\hat{\Omega}B')^{-1/2}B\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathcal{N}(0,I).$$

Thus, under these conditions, valid large sample confidence intervals and tests can be based on any variance estimator $\hat{\Omega}$ that is consistent in the sense that $\hat{\Omega}-\Omega_n \xrightarrow{p} 0$. We focus the theoretical discussion of this section on this consistency property.[2]

The asymptotic variance $\Omega_n$ described above depends on two matrices $\Gamma_n$ and $\Sigma_n$. Under the conditions of Theorem 1, $\hat{\Gamma}$ is a consistent estimator of $\Gamma_n$ even when $K/n$ is bounded away from zero, so we focus on estimation of $\Sigma_n$. An estimate can be formed by removing expectations and replacing disturbances by estimated residuals in the formula for $\Sigma_n$. For this approach to work the estimation error in the residuals must not be too large. For this purpose we use a potentially different choice of $K$ in forming the residuals than is used in the estimator $\beta$ or in the $M_{ij}$ that appears in $\Sigma_n$. Let $\tilde{K}$ denote this choice and define accordingly $\tilde{P} = [p^{\tilde{K}}(z_1), ..., p^{\tilde{K}}(z_n)]'$, $\tilde{Q} = \tilde{P}(\tilde{P}'\tilde{P})^{-}\tilde{P}'$, $\tilde{M} = I - \tilde{Q}$, $\breve{\varepsilon} = \tilde{M}(y - X\hat{\beta})$, and $\tilde{X} = \tilde{M}X$.

Therefore, for $\tilde{X} = [\tilde{x}_1, ..., \tilde{x}_n]'$ the variance estimator is given by

$$\hat{\Omega} = \hat{\Gamma}^{-1}\hat{\Sigma}\hat{\Gamma}^{-1}, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{1 \le i,j \le n} M_{ij}^2 \tilde{x}_i \tilde{x}_i' \breve{\varepsilon}_j^2,$$

where the $K$ used in $M_{ij}$ is the same as used for $\hat{\beta}$. The estimator $\hat{\Sigma}$ will be consistent for $\Sigma_n$ under appropriate regularity conditions. Those conditions will require that $\tilde{K}/n \to 0$, so that the estimation error in $\tilde{x}_i$ and $\breve{\varepsilon}_j$ is not too large. We do not yet have an estimator that is consistent without $\tilde{K}/n \to 0$.

The idea behind the construction of the estimator $\hat{\Sigma}$ parallels the intuitive discussion given in Section 2, which exploits the V-statistic representation of the estimator obtained under the generalized asymptotic experiment, and leads to a simple and "automatic" way of constructing asymptotically valid standard-errors under alternative asymptotics. This approach has also been implemented in other contexts, being implicitly based on the common structure of Section 3. For instance, for small bandwidth asymptotics this approach corresponds to the result in Theorem 2(a) of Cattaneo, Crump, and Jansson (2011) and for many instrument asymptotics to the result of Hansen, Hausman, and Newey (2008).

Under homoskedasticity of $\varepsilon_i$ the asymptotic variance simplifies and there is a corresponding

[2]Another approach to inference would be via the bootstrap. For small bandwidth asymptotics it has been shown by Cattaneo, Crump, and Jansson (2012) that the standard nonparametric bootstrap does not provide a valid distributional approximation in general. Similar results seem likely to hold for other alternative asymptotics, so we conjecture that the standard nonparametric bootstrap will not provide valid inference with many regressors.

relatively simple variance estimator. When $\mathbb{E}\left[\varepsilon_i^2|x_i, z_i\right] = \sigma_\varepsilon^2$ the asymptotic variance of $\hat{\beta}$ is $\sigma_\varepsilon^2\Gamma_n^{-1}$. As discussed earlier, $\hat{\Gamma}$ is an estimator of $\Gamma_n$, so we just need to estimate $\sigma_\varepsilon^2$ consistently. This can be done using the same $K$ as is used for $\hat{\beta}$ if a degrees of freedom correction is included. A consistent estimator of $\sigma_\varepsilon^2$ under many regressors is one with the degrees of freedom correction, leading to the homoskedasticity consistent asymptotic variance estimator

$$\hat{\Omega}_{HO} = s^2\hat{\Gamma}^{-1}, \qquad s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - d_x - K}, \qquad \hat{\varepsilon} = M(y - X\hat{\beta}).$$

We will show that this estimator is consistent under homoskedasticity, providing a many regressor justification for the degrees of freedom adjustment in $s^2$.

We will also study the properties of several other variance estimators that are available in the literature (for a review see, e.g., MacKinnon and White (1983), Chesher and Jewitt (1987), Cribari-Neto and Lima (2008) and references therein). Here we consider the best known and commonly used candidates. They belong to a class of variance estimators given by

$$\hat{\Omega}_k = \hat{\Gamma}^{-1}\hat{\Sigma}_k\hat{\Gamma}^{-1}, \qquad \hat{\Sigma}_k = X'M\hat{\Psi}_k MX/n, \qquad \hat{\Psi}_k = \begin{bmatrix} \omega_{k,1}\hat{\varepsilon}_1^2 & & 0 \\ & \ddots & \\ 0 & & \omega_{k,n}\hat{\varepsilon}_n^2 \end{bmatrix},$$

where $\omega_{k,i} = \omega_{k,i}(Z)$, $i = 1, ..., n$, are some appropriate weights that define the class of estimators. In particular, the choices $\omega_{1,i} = 1$, $\omega_{2,i} = n/(n - K - d_x)$, and $\omega_{3,i} = M_{ii}^{-\delta}$ for some $\delta \geq 1$ lead to well-known competing heteroskedasticity robust variance estimators commonly employed in the literature. We derive the properties of $\hat{\Omega}_k$ $(k = 1, 2, 3)$ with many regressors, finding that they are all inconsistent.

Turning now to the results, it is straightforward to show consistency of $s^2$, leading to consistency of $\hat{\Omega}_{HO}$ under homoskedasticity.

**Theorem 2** If the hypotheses of Theorem 1 are satisfied, $\mathbb{E}\left[\varepsilon_i^2|x_i, z_i\right] = \sigma_\varepsilon^2 > 0$, and $P'P$ is nonsingular a.s., then $s^2 \xrightarrow{p} \sigma_\varepsilon^2$ and $\hat{\Omega}_{HO} - \Omega_n = s^2\hat{\Gamma}^{-1} - \sigma_\varepsilon^2\Gamma_n^{-1} \xrightarrow{p} 0$.

This result provides a distribution free, large sample justification for the degrees-of-freedom correction required for exact inference under homoskedastic Gaussian errors. Intuitively, accounting for the correct degrees of freedom is important whenever the number of terms in the linear model is "large" relative to the sample size.

Turning now to properties of the other estimators, we give a general result that includes consistency of $\hat{\Omega}$ and the limits of $\hat{\Omega}_k$, $k = 1, 2, 3$ and $\delta \geq 1$. Recall that the definitions of $M$ and $\tilde{M}$ given previously. Let $\omega_{ij}$ denote a nonnegative weight for each $i$ and $j$ and

$$\hat{\Upsilon} = \frac{1}{n} \sum_{1 \leq i,j \leq n} \omega_{ij} \tilde{x}_i \tilde{x}_i' \breve{\varepsilon}_j^2.$$

Note that for $\omega_{ij} = M_{ij}^2$ we have $\hat{\Upsilon} = \hat{\Omega}$, while for $\tilde{K} = K$ and $\omega_{ij} = 1(i = j)\omega_{i,k}$, $k = 1, 2, 3$ and some $\delta \geq 1$, we will have $\hat{\Upsilon} = \hat{\Omega}_k$. Thus, a general result on the limit of this object can be used to find the limit of the estimators we have described.

**Lemma 1.** If the hypotheses of Theorem 1 are satisfied, $\sqrt{n}\tilde{K}^{-\alpha_g-\alpha_h} \to 0$, $\omega_{ij} = \omega_{ij}(Z)$ is positive, symmetric, bounded, and there is a constant $C$ such that $\sum_{j=1}^{n} \omega_{ij} \leq C$ (a.s.) for all $n$ and $i$ with $1 \leq i \leq n$, then,

$$\hat{\Upsilon} = \bar{\Upsilon}_n + o_p(1), \qquad \bar{\Upsilon}_n = \frac{1}{n} \sum_{1 \leq i,j \leq n} \left( \sum_{1 \leq \ell_1, \ell_2 \leq n} \omega_{\ell_1 \ell_2} \tilde{M}_{i\ell_1}^2 \tilde{M}_{j\ell_2}^2 \right) \mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j].$$

Consistency of $\hat{\Omega}$ in the general heteroskedastic case follows from this result.

**Theorem 3.** If the hypotheses of Lemma 1 are satisfied and $\tilde{K}/n \to 0$ then $\hat{\Omega} - \Omega_n \xrightarrow{p} 0$.

Lemma 1 can also be specialized to show that the Eicker-White heteroskedasticity robust variance estimator is consistent if $K/n \to 0$, under the Donald and Newey (1994) conditions. We report this result here because it is apparently new, though our main focus is on cases where $K/n$ does not go to zero.

**Theorem 4.** If the hypotheses of Theorem 1 are satisfied and $K/n \to 0$, then $\hat{\Sigma}_k = \Sigma_n + o_p(1)$, $k = 1, 2$.

Under homoskedasticity we can make explicit comparisons between the standard heteroskedasticity estimators and the asymptotic variance $\Sigma_n$.

**Theorem 5.** If the hypotheses of Theorem 1 are satisfied, $\tilde{K} = K$, and $\mathbb{E}[\varepsilon_i^2 | x_i, z_i] = \sigma_\varepsilon^2$ then

$$\hat{\Sigma}_1 = \bar{\Sigma}_{1,n} + o_p(1), \qquad \bar{\Sigma}_{1,n} = \Sigma_n - \frac{\sigma_\varepsilon^2}{n} \sum_{1 \le i,j \le n} M_{ij}^2 \left(1 - M_{jj}\right) \mathbb{E}[v_i v_i' | z_i] < \Sigma_n.$$

Also, if $Q_{ii} \geq C > 0$ then

$$\hat{\Sigma}_3 = \begin{cases} \Sigma_n + o_p(1) & \text{if } \delta = 1 \\ \bar{\Sigma}_n(\delta) + o_p(1), \qquad \bar{\Sigma}_n(\delta) > \Sigma_n & \text{if } \delta > 1 \end{cases}.$$

Thus we find that the Eicker-White variance estimator $\hat{\Omega}_1$ is downward biased under homoskedasticity when $K/n$ does not go to zero. In contrast, for $k = 3$ with $\delta = 1$ (i.e., $\omega_{3,i} = M_{ii}^{-1}$) the resulting estimator $\hat{\Omega}_3$, which is approximately unbiased under homoskedasticity, is consistent even when $K/n$ does not go to zero, provided that $\varepsilon_i$ is in fact homoskedastic. This result provides a large sample justification for the use of $\hat{\Omega}_3$ when many regressors are employed in a homoskedastic linear model. However, for $k = 3$ with $\delta > 1$ (i.e., $\omega_{3,i} = M_{ii}^{-\delta}$) the resulting estimator $\hat{\Omega}_3$ is upward biased under homoskedasticity.

The previous discussion shows that when $\varepsilon_i$ are homoskedastic all but one of the commonly used heteroskedasticity-robust standard-error estimators will be inconsistent unless $K/n \to 0$. The only estimator that remains consistent when $K/n \to \alpha \in [0, 1)$ in the homoskedastic case is $\hat{\Omega}_3$ with the specific choice $\delta = 1$, which is specially designed to be (approximately) unbiased under homoskedasticity. However, when $\varepsilon_i$ are in fact heteroskedastic, it is possible to show by example that $\hat{\Omega}_3$ with $\delta = 1$ will also be inconsistent in general. For instance, if $d_x = 1$ and $\mathbb{E}[v_i^2 \varepsilon_j^2 | Z] = \mathbb{E}[v_i^2 | Z] \mathbb{E}[\varepsilon_j^2 | Z] = c_i c_j$ for all $1 \le i, j \le n$, then it is not difficult to show that $\hat{\Sigma}_3 = \Sigma_n + \lambda' C_n \lambda + o_p(1)$ where $\lambda \in \mathbb{R}_{++}^{d_x}$ and $C_n$ is a non-zero matrix in general.

In summary, none of standard heteroskedasticity robust variance estimators are consistent when $K/n$ does not go to zero. Thus, there exists an important sense in which the classical heteroskedasticity-robust standard-error estimators typically employed in the literature are not robust, even in a simple linear model, provided that the number of regressors is large relative to the sample size. The failure of these estimators is due to the fact that both $\tilde{x}_i$ and $\hat{\varepsilon}_i$ are estimated with too much variability when $K$ is large relative to $n$. (Recall that consistency of nonparametric series estimators requires $K/n \to 0$; see, e.g., Newey (1997).)

## 5 Small Simulation Study

We conducted an small-scale Monte Carlo experiment to explore the extent to which the asymptotic theoretical results obtained in the previous sections are present in small samples. We consider the following simple partially linear model:

$$y_i = x_i'\beta + g(z_i) + \varepsilon_i, \qquad \varepsilon_i = \sigma_\varepsilon(x_i, z_i)\, u_{1i},$$
$$x_i = h(z_i) + v_i, \qquad v_i = \sigma_v(z_i)\, u_{2i},$$

where $d_x = 1$, $d_z = 10$, $\beta = 0$, $z_i = (z_{1i}, ..., z_{d_z i})'$ with $z_{\ell i} \sim \mathcal{U}(-1, 1)$, $\ell = 1, ..., d_z$, $u_{1i} \sim \mathcal{N}(0, 1)$ and $u_{2i} \sim \mathcal{U}(-1, 1)$. The simulation study is based on $S = 10{,}000$ replications, each replication taking a random sample of size $n = 300$ with all random variables generated independently. For simplicity, the functional forms of the regression functions are chosen to be additive separable and equal, $g(z_i) = 1 + g_1(z_{1i}) + ... + g_{10}(z_{10i})$ and $h(z_i) = 1 + h_1(z_{1i}) + ... + h_{10}(z_{10i})$, with $g_\ell(z_{\ell i}) = h_\ell(z_{\ell i}) = z_{\ell i}(2 + z_{\ell i})^{-1/2}$ for $\ell = 1, ..., d_z$. We consider two data generating processes (DGP) depending on the form of heteroskedasticity imposed. Model 1 sets $\sigma^2_\varepsilon(x_i, z_i) = \sigma^2_v(z_i) = 1$, and therefore corresponds to a homoskedastic DGP, while Model 2 sets $\sigma^2_\varepsilon(x_i, z_i) = (z_i'\iota + x_i)^6/10^4$ and $\sigma^2_v(z_i) = (z_i'\iota)^6/10^4$, with $\iota = (1, 1, \cdots, 1)' \in \mathbb{R}^{d_z}$, giving a heteroskedastic DGP.

The estimators considered in the Monte Carlo experiment exploit the known additive separable structure of $g(z_i)$ and $h(z_i)$, and are based on power series. Specifically, we approximate $g_\ell(z_{\ell i})$ by $p^K(z_{\ell i})'\gamma_\ell$, $\ell = 1, 2, \cdots, d_z$, with $p^K(z_{\ell i}) = (1, z_{\ell i}, z^2_{\ell i}, \cdots, z^K_{\ell i})'$. To explore the consequences of introducing many regressors in the partially linear model, we focus on the empirical coverage rate of the 8 competing confidence intervals introduced in Section 5 for a grid of $K = 0, 1, ..., 10$. (Note that $K = 0$ corresponds to including only a constant term, and then each increase in $K$ corresponds to including 10 more regressors.)

We focus on coverage probabilities for confidence intervals constructed from asymptotic t-ratios, each taking the form $T = (\hat{\beta} - \beta)/\sqrt{\tilde{\Omega}/n}$, where $\tilde{\Omega}$ is one of six possible estimators. The interval for $T_{AL}$ corresponds to the consistent variance estimator, $\tilde{\Omega} = \hat{\Omega}$, using $\tilde{K} = \min\{K, K_{CV}\}$, where $K_{CV}$ corresponds to the $K$ chosen by standard cross-validation on the residuals of the partially linear model. The intervals corresponding to $T_{HO,k}$ $(k = 1, 2)$ are obtained using the $\tilde{\Omega} = \hat{\Omega}_{HO}$ for $k = 2$ and the variance estimator without a degrees of freedom correction $\tilde{\Omega} = (n - 1 - K)\hat{\Omega}_{HO}/n$ for $k = 1$. The intervals for $T_{HE,k}$ correspond to $\hat{\Omega}_k$, with $k = 1, 2, 3$ and $\delta = 1$.

The main findings from the Monte Carlo experiment are presented in Tables 1. We explored many other specifications for the regression functions, heteroskedasticity form, distributional assumptions, basis of approximation, etc., but we do not include these additional results because they were qualitative similar to those reported here. In general, all the results are consistent with the theoretical conclusions presented in the previous sections. First, we found that the results for Gaussian and non-Gaussian errors are qualitatively similar. Second, in most cases a small choice of $K$ leads to important biases that affect the empirical size of all the confidence intervals. Third, under homoskedasticity, as $K$ becomes larger relative to the sample size confidence intervals without degrees-of-freedom correction ($T_{HO,1}$ and $T_{HE,1}$) are under-sized, while the analogue confidence intervals with degrees-of-freedom correction ($T_{HO,2}$ and $T_{HE,2}$) have close-to-correct empirical size. In addition, $T_{HE,3}$ (approximately unbiased under homoskedasticity) has also good empirical size properties. The new confidence intervals based on $T_{AL}$ perform well under homoskedasticity. Fourth, under heteroskedasticity, we found that $T_{HO,1}$, $T_{HO,2}$, $T_{HE,1}$, $T_{HE,2}$ and $T_{HE,3}$ were all under-sized as the number of regressors grows relative to the sample size. This result shows that neither degrees-of-freedom correction nor employing the heteroskedasticity-robust standard-error estimator which is approximately unbiased under homoskedasticity lead to close-to-correct empirical size when $K/n$ is "large". In contrast, the new confidence intervals introduced in this paper ($T_{ALT}$) exhibited good empirical coverage for the full range of $K/n$ considered in the simulation study, once bias became small, under heteroskedasticity.

Table 1: Empirical Coverage Rates of 95% Confidence Intervals

| | Model 1 (Homoskedasticity) | | | | | | Model 2 (Heteroskedasticity) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $K/n$ | $T_{HO,1}$ | $T_{HO,2}$ | $T_{HE,1}$ | $T_{HE,2}$ | $T_{HE,3}$ | $T_{AL}$ | $T_{HO,1}$ | $T_{HO,2}$ | $T_{HE,1}$ | $T_{HE,2}$ | $T_{HE,3}$ | $T_{AL}$ |
| 0.003 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 57.3 | 57.5 | 79.4 | 79.6 | 79.5 | 79.3 |
| 0.037 | 79.7 | 80.8 | 79.1 | 80.0 | 79.8 | 78.2 | 83.3 | 84.2 | 93.8 | 94.3 | 94.3 | 98.5 |
| 0.070 | 93.5 | 94.4 | 93.4 | 94.4 | 94.3 | 93.8 | 64.6 | 66.2 | 90.1 | 91.4 | 91.5 | 97.1 |
| 0.103 | 93.1 | 94.6 | 92.9 | 94.4 | 94.3 | 93.8 | 62.6 | 65.3 | 88.5 | 90.6 | 90.8 | 97.0 |
| 0.137 | 92.8 | 94.8 | 92.6 | 94.5 | 94.5 | 94.0 | 62.6 | 66.1 | 87.1 | 90.0 | 90.3 | 97.0 |
| 0.170 | 92.3 | 94.5 | 92.0 | 94.2 | 94.2 | 93.9 | 62.5 | 66.8 | 85.3 | 89.4 | 89.7 | 97.0 |
| 0.203 | 91.7 | 94.6 | 91.6 | 94.5 | 94.4 | 93.9 | 62.2 | 67.6 | 84.0 | 88.8 | 89.2 | 97.1 |
| 0.237 | 91.0 | 94.7 | 90.7 | 94.4 | 94.2 | 93.9 | 62.0 | 68.5 | 82.3 | 88.3 | 88.7 | 97.1 |
| 0.270 | 90.5 | 94.8 | 90.2 | 94.7 | 94.6 | 94.3 | 61.6 | 68.6 | 80.7 | 88.0 | 88.4 | 97.2 |
| 0.303 | 89.6 | 94.9 | 89.6 | 94.8 | 94.6 | 94.4 | 61.2 | 69.3 | 78.8 | 87.7 | 88.2 | 97.2 |
| 0.337 | 88.8 | 95.0 | 88.6 | 94.8 | 94.7 | 94.3 | 60.8 | 70.4 | 77.2 | 87.2 | 87.6 | 97.3 |

In conclusion, we found in our small-scale simulation study that $T_{AL}$ leads to confidence intervals exhibiting good empirical coverage under both homoskedasticity and heteroskedasticity even when $K/n$ is "large". None of the standard competing confidence intervals considered here appear to have a similar property. These findings are consistent with the theoretical results obtained in the paper.

# 6 Conclusion

This paper showed that the many instrument asymptotics and the small bandwidth asymptotics shared a common structure based on a V-statistic, with a remainder term that is asymptotically normal when the number of term diverges to infinity or the bandwidth shrinks to zero. This feature is particularly useful to obtain new results for other semiparametric estimators. In this paper we employ this common structure to derive a new alternative large-sample distributional approximation for a series estimator of the partially linear model. This result not only implied a new (larger) asymptotic variance formula, but also led to a detailed analysis of commonly used standard-error estimators for linear models and the development of a new standard-error formula that is consistent under both heteroskedasticity of unknown form and many terms asymptotics.

## References

Angrist, J., G. W. Imbens, and A. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14(1), 57–67.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, 62, 657–681.

Calhoun, G. (2011): "Hypothesis Testing in Linear Regression when $k/n$ is large," *Journal of Econometrics*, 165, 163–174.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2010): "Robust Data-Driven Inference for Density-Weighted Average Derivatives," *Journal of the American Statistical Association*, 105(491), 1070–1083.

——— (2011): "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," forthcoming in *Econometric Theory*.

——— (2012): "Bootstrapping Density-Weighted Average Derivatives," Working paper, University of Michigan.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): "Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments," forthcoming in *Econometric Theory*.

Chesher, A., and I. Jewitt (1987): "The Bias of a heteroskedasticity consistent covariance matrix estimator," *Econometrica*, 55(5), 1217–1222.

Cribari-Neto, F., and M. Lima (2008): "Sequences of bias-adjusted covariance matrix estimators under heteroskedasticity of unknown form," *Annals of the Institute of Statistical Mathematics*, 62(6), 1053–1082.

Donald, S. G., and W. K. Newey (1994): "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50(1), 30–40.

Hansen, C., J. Hausman, and W. K. Newey (2008): "Estimation with Many Instrumental Variables," *Journal of Business and Economic Statistics*, 26(4), 398–422.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Kolesar, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens (2011): "Identification and Inference with Many Invalid Instruments," working paper, Harvard University.

Kunitomo, N. (1980): "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, 75(371), 693–700.

Linton, O. (1995): "Second Order Approximation in the Partialy Linear Regression Model," *Econometrica*, 63(5), 1079–1112.

MacKinnon, J., and H. White (1983): "Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305–325.

Morimune, K. (1983): "Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size," *Econometrica*, 51(3), 821–841.

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

Newey, W. K., F. Hsieh, and J. M. Robins (2004): "Twicing Kernels and a Small Bias Property of Semiparametric Estimators," *Econometrica*, 72, 947–962.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Stock, J. H., and M. W. Watson (2007): "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression," *Econometrica*, 76(1), 155–174.

van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, New York.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

# Supplemental Appendix to "Alternative Asymptotics and the Partially Linear Model with Many Regressors"

Matias D. Cattaneo, Department of Economics, University of Michigan.

Michael Jansson, Department of Economics, UC Berkeley and CREATES.

Whitney K. Newey, Department of Economics, MIT.

January, 2012

## 1 Derivation of Examples

**Example 1: "Many Instrument Asymptotics".** Consider a linear structural equation

$$y_i = x_i'\beta_0 + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i] = 0, \quad i = 1, ..., n,$$

where $x_i$ is a vector and $y_i$ and $\varepsilon_i$ are scalar dependent variable and disturbance respectively. Let $z_i$ be a $K \times 1$ vector of instrumental variables that we treat as constants. It is also equivalent to allow instrumental variables to be random but condition on the matrix of observations $Z = [z_1, ..., z_n]'$ and replace unconditional moment conditions with conditional ones. To describe the estimator let $Q = Z(Z'Z)^{-}Z'$ denote the projection matrix on the column space of $Z$. After centering and scaling, the JIVE2 statistic takes the form

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left( \sum_{1 \leq i,j \leq n, i \neq j} Q_{ij} x_i x_j'/n \right)^{-1} \sum_{1 \leq i,j \leq n, i \neq j} Q_{ij} x_i \varepsilon_j / \sqrt{n},$$

where $Q_{ij}$ represents the $(i,j)$-th element of the matrix $Q$. Therefore, JIVE2 is a special case of equation (1) with

$$\mu_0 = \beta_0, \quad \hat{\Gamma} = \sum_{1 \leq i,j \leq n, i \neq j} Q_{ij} x_i x_j'/n,$$

$$u_{ii}^n(W_i, W_i) = 0, \quad u_{ij}^n(W_i, W_j) = Q_{ij} x_i \varepsilon_j / \sqrt{n} \quad \text{for } i \neq j.$$

Note that $\mathbb{E}[u_{ij}^n(W_i, W_j)] = 0$ and for $\Upsilon_i = \mathbb{E}[x_i]$,

$$\mathbb{E}[u_{ij}^n(W_i, W_j)|W_i] = Q_{ij} x_i \mathbb{E}[\varepsilon_j] = 0, \quad \mathbb{E}[u_{ij}^n(W_i, W_j)|W_j] = Q_{ij} \Upsilon_i \varepsilon_j / \sqrt{n}.$$

Here $\Upsilon_i$ can be interpreted as the reduced form for observation $i$. Let $v_i = x_i - \Upsilon_i$. Applying Proposition 1, and using symmetry of the matrix $Q$, we find that equation (2) is satisfied with

$$\psi_i^n(W_i) = \left( \sum_{1 \le i, j \le n, i \neq j} Q_{ij} \Upsilon_j \right) \varepsilon_i = \Upsilon_i (1 - Q_{ii}) \varepsilon_i / \sqrt{n} - \left( \Upsilon_i - \sum_{1 \le i \le n} Q_{ij} \Upsilon_j \right) \varepsilon_i / \sqrt{n},$$

$$D_i^n(W_i, ..., W_1) = \sum_{1 \le j \le n, j < i} Q_{ij} \left( v_i \varepsilon_j + v_j \varepsilon_i \right) / \sqrt{n}, \quad B_n = 0.$$

Note that $\Upsilon_i - \sum_{j=1}^n Q_{ij} \Upsilon_j$ is the $i$-th residual from regressing the reduced form observations on $Z$, so that by appropriate definition of the reduced form this can generally be assumed to vanish as the sample size grows. In that case

$$\Psi_n = \sum_{1 \le i \le n} \Upsilon_i (1 - Q_{ii}) \varepsilon_i / \sqrt{n} + o_p(1).$$

Furthermore, under standard asymptotics $Q_{ii}$ will go to zero, so the variance of this does indeed correspond to the usual asymptotic variance for IV.

The degenerate U-statistic term is

$$U_n = \sum_{1 \le j < i \le n} Q_{ij} \left( v_i \varepsilon_j + v_j \varepsilon_i \right) / \sqrt{n}.$$

Chao, Swanson, Hausman, Newey, and Woutersen (2012) apply the martingale central limit theorem to show that this $U_n$ will be asymptotically normal when $x_i$ and $\varepsilon_i$ have uniformly bounded fourth moments, $\text{rank}(Q) = \dim(Z) = K \to \infty$, and $Q_{ii}$ is bounded away from 1 uniformly in $i$ and $n$. The conditions of the martingale central limit theorem are verified by showing that certain linear combinations with coefficients depending on the elements of $Q$ go to zero as $K \to \infty$. In the proof, this makes individual terms asymptotically negligible, with a Lindeberg-Feller condition being satisfied. Alternative asymptotics occurs when $K$ grows as fast as $n$, resulting in $\Psi_n$ and $U_n$ having the same magnitude in the limit. ■

**Example 2: "Small Bandwidth Asymptotics".** Consider the kernel estimation of the integrated squared density given by $\mu_0 = \int f_0(w)^2 \mathrm{d}w = \mathbb{E}[f_0(W_i)]$, where $W_i$ denotes a continuously distributed random variable with p.d.f. $f_0$. A leave-one-out estimator is

$$\hat{\mu} = \sum_{1 \le i,j \le n, i \ne j} K_h(W_i - W_j)/n(n-1),$$

where $K(u)$ is a symmetric kernel and $K_h(u) = h^{-d} K(u/h)$. This estimator has the V-statistic form of equation (1) with $\mu_0 = \int f_0(w)^2 \mathrm{d}w$, $\hat{\Gamma} = 1$,

$$u_{ii}^n(W_i, W_i) = 0, \quad u_{ij}^n(W_i, W_j) = K_h(W_i - W_j)/\sqrt{n}(n-1) \quad \text{for } i \ne j.$$

Define $f_h(w) = \int K(u) f(w + hu) \mathrm{d}u$. Note that by symmetry of $K(u)$,

$$\mathbb{E}[u_{ij}^n(W_i, W_j)|W_i] = f_h(W_i)/\sqrt{n}(n-1), \quad \mathbb{E}[u_{ij}^n(W_i, W_j)|W_j] = f_h(W_j)/\sqrt{n}(n-1),$$

$$\mathbb{E}[u_{ij}^n(W_i, W_j)] = \mathbb{E}[f_h(X_i)]/\sqrt{n}(n-1).$$

Applying Proposition 1, and using symmetry of $K(u)$, we find that equation (2) is satisfied with

$$\psi_i^n(W_i) = 2\{f_h(W_i) - \mathbb{E}[f_h(W_i)]\}/\sqrt{n},$$

$$D_i^n(W_i, ..., W_1) = 2 \sum_{1 \le j \le n, j < i} \{K_h(W_i - W_j) - f_h(W_i) - f_h(W_j) + \mathbb{E}[f_h(W_i)]\}/\sqrt{n}(n-1),$$

$$B_n = \sqrt{n}\{\mathbb{E}[f_h(W_i)] - \mu_0\}.$$

Note that $2\{f_h(W_i) - \mathbb{E}[f_h(W_i)]\}$ is an approximation to the well known influence function $2[f_0(W_i) - \mu_0]$ for estimators of the integrated squared density. As $h \to 0$ we will have $f_h(W_i)$ converging to $f_0(W_i)$ in mean square, so that

$$\Psi_n = \sum_{1 \leq i \leq n} 2[f_0(W_i) - \mu_0]/\sqrt{n} + o_p(1).$$

The asymptotic variance of this leading term does correspond to the usual asymptotic variance for estimators of the integrated square density.

The degenerate U-statistic term is

$$U_n = 2 \sum_{1 \leq i < j \leq n} \{K_h(W_i - W_j) - f_h(W_i) - f_h(W_j) + \mathbb{E}[f_h(W_i)]\}/\sqrt{n}(n-1).$$

The martingale central limit theorem can be applied as in Cattaneo, Crump, and Jansson (2011) to show that this $U_n$ will be asymptotically normal as $h \to 0$ and $n \to \infty$ growing, provided that $n^2 h^d \to \infty$. It is easy to show that $n^2 h^d \mathbb{V}[U_n] \to \Delta = \mu_0 \int K(u)^2 \mathrm{d}u$ under appropriate moment and smoothness conditions. Intuitively, the bandwidth shrinking leads to $K_h(W_i - W_j)$ being small except when $W_i$ and $W_j$ are close, so that tail of the distribution of $\sqrt{n} D_i^n(W_i, ..., W_1)$ becomes thin and a Lindeberg-Feller condition is satisfied. Alternative asymptotics occurs when $h^d$ shrinks as fast as $1/n$, resulting in $\Psi_n$ and $U_n$ having the same magnitude in the limit. ■

## 2 Proofs of Results

All statements involving conditional expectations are understood to hold almost surely. Qualifiers such as "a.s." will be omitted to conserve space. Throughout the appendix, $C$ will denote a generic constant that may take different values in each case. Recall that $M = I - Q$ is symmetric and idempotent, and therefore $|M_{ij}| \leq M_{ii} \leq 1$, $n - K = \sum_{1 \leq i \leq n} M_{ii}$ and $M_{ij} = \sum_{1 \leq \ell \leq n} M_{i\ell} M_{\ell j}$.

To prove the main results of the paper we will employ the following three preliminary lemmas. The first lemma provides a formal justification for the asymptotic expansion of $\hat{\Gamma}$.

**Lemma A-1.** (Matrix $\hat{\Gamma}$) If Assumptions 1(a) and 2 are satisfied then

$$\hat{\Gamma} = \Gamma_n + o_p(K/n), \qquad \Gamma_n = \frac{1}{n}\sum_{1\le i\le n} M_{ii}\mathbb{E}[v_i v_i'|z_i].$$

**Proof of Lemma A-1**. By Assumption 1(a), $MP = 0$, and the Markov inequality we have $\text{trace}(H'MH/n) = (H - P\eta_h')'M(H - P\eta_h')/n \le \sum_{1\le i\le n} \|h(z_i) - \eta_h p^K(z_i)\|^2/n = O_p(K^{-2\alpha_h}) \xrightarrow{p} 0$. Also, by Lemma A1 of Chao, Swanson, Hausman, Newey, and Woutersen (2012) and $\sqrt{K}/n \to 0$ it follows that $\sum_{1\le i,j\le n, j\ne i} Q_{ij} v_i v_j'/n \xrightarrow{p} 0$. By independence, $v_i v_i'$ and $v_j v_j'$ are uncorrelated conditional on $Z$. By $M_{ii} \le 1$ and Assumption 2 we also have $\mathbb{E}[M_{ii}^2\|v_i\|^4|Z] \le \mathbb{E}[\|v_i\|^4|Z] \le C$. Therefore, by Chebyshev inequality,

$$\sum_{1\le i\le n} M_{ii} v_i v_i'/n - \mathbb{E}\left[\sum_{1\le i\le n} M_{ii} v_i v_i'/n \middle| Z\right] = \sum_{1\le i\le n} M_{ii} v_i v_i'/n - \Gamma_n \xrightarrow{p} 0.$$

Also, $V'V/n = O_p(1)$ by the Markov inequality, so by the Cauchy-Schwartz inequality and $M$ idempotent, $\|H'MV/n\| \le [\text{trace}(H'MH/n)\text{trace}(V'V/n)]^{1/2} \xrightarrow{p} 0$. By the triangle inequality we then have $X'MX/n = (V+H)'M(V+H)/n = V'MV/n + o_p(1)$ with

$$V'MV/n = \sum_{1\le i\le n} M_{ii} v_i v_i'/n - \sum_{1\le i,j\le n, j\ne i} Q_{ij} v_i v_j'/n + o_p(1) = \Gamma_n + o_p(1),$$

which gives the result. Q.E.D.

The next lemma establishes the order of the different "biases" present in the statistics $S_n = X'M(\varepsilon + G)$.

**Lemma A-2.** (Approximation Bias)

(a) If Assumption 1 is satisfied then $B_n = H'MG/\sqrt{n} = O_p(\sqrt{n}K^{-(\alpha_h+\alpha_g)})$.

(b) If Assumptions 1–2 are satisfied then $(V'MG + H'M\varepsilon)/\sqrt{n} = O_p(K^{-\alpha_h} + K^{-\alpha_g})$.

**Proof of Lemma A-2**. By the Cauchy-Schwartz inequality and $M$ idempotent, $\|H'MG/n\| \le [\text{trace}(H'MH/n)\text{trace}(G'MG/n)]^{1/2} = O_p(K^{-\alpha_h})O_p(K^{-\alpha_g})$, which gives part (a) of the lemma.

For part (b), observe that $\mathbb{E}[\|H'M\varepsilon/\sqrt{n}\|^2|Z] = \text{trace}(H'M\mathbb{E}[\varepsilon\varepsilon'|Z]MH)/n \leq C\text{trace}(H'MH)/n = O_p(K^{-2\alpha_h})$, and $\mathbb{E}[\|V'MG/\sqrt{n}\|^2|Z] = G'M\mathbb{E}[VV'|Z]MG/n \leq CG'MG/n = O_p(K^{-2\alpha_g})$, which gives the result by Markov inequality. Q.E.D.

Finally, the next lemma establishes a valid Gaussian distributional approximation for an appropriately standardized statistic, even when $K/n$ is bounded away from zero.

**Lemma A-3.** If Assumption 2 is satisfied, $K \to \infty$, and there is $0 < C < 1$ such that $Q_{ii} < C$, $\lambda_{\min}(\Sigma_n) \geq 1/C$ then,

$$\Sigma_n^{-1/2} V'M\varepsilon/\sqrt{n} \xrightarrow{d} \mathcal{N}(0, I_{d_x}).$$

**Proof of Lemma A-3**. Follows by Lemma A2 of Chao, Swanson, Hausman, Newey, and Woutersen (2012). Q.E.D.

Next, we turn to the proof of the results stated in the main text.

**Proof of Theorem 1**. The proof follows by straightforward algebra, together with the results from Lemmas A1–A3 and applying the Slutsky Theorem. Q.E.D.

**Proof of Theorem 2**. First, it follows from $G'MG/n = o_p(1)$ and the Cauchy-Schwarz inequality that $\hat{\varepsilon}'\hat{\varepsilon}/n = (Y - X\hat{\beta} - G)'M(Y - X\hat{\beta} - G)/n + o_p(1)$, provided $(Y - X\hat{\beta} - G)'M(Y - X\hat{\beta} - G)/n = O_p(1)$. Next, note that Lemma A-1 and $\hat{\beta} - \beta = o_p(1)$ imply $(\hat{\beta} - \beta)'X'MX(\hat{\beta} - \beta)/n = o_p(1)$, which together with the Cauchy-Schwarz inequality gives $(Y - X\hat{\beta} - G)'M(Y - X\hat{\beta} - G)/n = \varepsilon'M\varepsilon/n + (\hat{\beta} - \beta)'X'MX(\hat{\beta} - \beta)/n - 2(Y - X(\hat{\beta} - \beta) - G)'MX(\hat{\beta} - \beta)/n = \varepsilon'M\varepsilon/n + o_p(1)$. Finally, it follows similarly to the proof of Lemma A-1 that

$$\varepsilon'M\varepsilon/n = \sum_{1\leq i\leq n} M_{ii}\varepsilon_i^2/n - \sum_{1\leq i,j\leq n, j\neq i} \varepsilon_i Q_{ij}\varepsilon_j/n = \sum_{1\leq i\leq n} M_{ii}\mathbb{E}[\varepsilon_i^2|Z]/n + o_p(1) = \frac{n-K}{n}\sigma_\varepsilon^2 + o_p(1).$$

The conclusion follows by the triangle inequality. Q.E.D.

**Proof of Lemma 4**. Note that $\breve{\varepsilon} = \tilde{Y} - \tilde{X}'\hat{\beta} = \tilde{\varepsilon} - \tilde{X}'(\hat{\beta} - \beta) + \tilde{G}$, for $\tilde{Y} = \tilde{M}Y$, $\tilde{X} = \tilde{M}X$,

$\tilde{\varepsilon} = \tilde{M}\varepsilon$, and $\tilde{G} = \tilde{M}G$. Thus, $\hat{\Upsilon}_n = \tilde{\Upsilon}_n + \epsilon_{1,n} + \epsilon_{2,n} + \epsilon_{3,n} + \epsilon_{4,n} - 2\epsilon_{5,n}$, where

$$\tilde{\Upsilon}_n = \sum_{1\le i,j\le n} \omega_{ij}\tilde{\varepsilon}_i^2 \tilde{v}_j \tilde{v}_j'/n, \quad \epsilon_{1,n} = \sum_{1\le i,j\le n} \omega_{ij}\tilde{\varepsilon}_i^2 \tilde{v}_j \left(\tilde{x}_j - \tilde{v}_j\right)'/n,$$

$$\epsilon_{2,n} = \sum_{1\le i,j\le n} \omega_{ij}\tilde{\varepsilon}_i^2 \left(\tilde{x}_j - \tilde{v}_j\right)\tilde{v}_j'/n, \quad \epsilon_{3,n} = \sum_{1\le i,j\le n} \omega_{ij}\tilde{\varepsilon}_i^2 \left(\tilde{x}_j - \tilde{v}_j\right)\left(\tilde{x}_j - \tilde{v}_j\right)'/n,$$

$$\epsilon_{4,n} = \sum_{1\le i,j\le n} \omega_{ij}(\tilde{x}_i'(\hat{\beta}-\beta) + \tilde{g}_i)^2 \tilde{x}_j\tilde{x}_j'/n, \quad \epsilon_{5,n} = \sum_{1\le i,j\le n} \omega_{ij}\tilde{\varepsilon}_i(\tilde{x}_i'(\hat{\beta}-\beta) + \tilde{g}_i)\tilde{x}_j\tilde{x}_j'/n,$$

with $\tilde{X} = [\tilde{x}_1, \cdots, \tilde{x}_n]'$, $\tilde{V} = \tilde{M}V = [\tilde{v}_1, \cdots, \tilde{v}_n]'$, and $\tilde{G} = (\tilde{g}_1, ..., \tilde{g}_n)'$. Note that that $\tilde{\varepsilon}_i = \varepsilon_i - \sum_{1\le j\le n} \tilde{Q}_{ij}\varepsilon_j$ and $\tilde{v}_i = v_i - \sum_{1\le j\le n} \tilde{Q}_{ij} v_j$. Since $\mathbb{E}[\varepsilon_i|z_i] = 0$ and $\mathbb{E}[v_i|z_i] = 0$, $\mathbb{E}[\tilde{\varepsilon}_i^4|Z] \le C\sum_{1\le j,k\le n} \tilde{M}_{ij}^2\tilde{M}_{ik}^2\mathbb{E}[\varepsilon_j^2\varepsilon_k^2|z_j, z_k] \le C$ and $\mathbb{E}[\|\tilde{v}_i\|^4|Z] \le C$ by properties of the idempotent matrixes. Also, for $\tilde{H} = [\tilde{h}_1, ..., \tilde{h}_n]'$,

$$\begin{aligned}
\sum_{1\le i\le n} \tilde{g}_i^2 &= G'\tilde{M}G \le \sum_{1\le i\le n} (g(z_i) - p^{\tilde{K}}(z_i)'\eta_g)^2 = O_{as}(n\tilde{K}^{-2\alpha_g}), \\
\sum_{1\le i\le n} \|\tilde{h}_i\|^2 &= \text{trace}(H'\tilde{M}H) \le \sum_{1\le i\le n} \|h(z_i) - p^{\tilde{K}}(z_i)'\eta_h\|^2 = O_{as}(n\tilde{K}^{-2\alpha_h})
\end{aligned}$$

Using $\tilde{x}_i - \tilde{v}_i = \tilde{h}_i$, it follows that $\epsilon_{1,n} = o_p(1)$ because the Cauchy-Schwarz inequality gives

$$\begin{aligned}
\mathbb{E}[\|\epsilon_{1,n}\||Z] &\le C\sum_{1\le i,j\le n} \omega_{ij}\mathbb{E}[\tilde{\varepsilon}_i^2\|\tilde{v}_j\||Z]\|\tilde{h}_i\|/n \\
&\le C\sum_{1\le i\le n} \|\tilde{h}_i\|/n \le C\sqrt{\sum_{1\le j\le n} \|\tilde{h}_i\|^2/n} = O_{as}(\tilde{K}^{-\alpha_h}),
\end{aligned}$$

and $\epsilon_{2,n} = O_{as}(\tilde{K}^{-\alpha_h})$ by the same argument. Similarly, $\epsilon_{3,n} = O_{as}(\tilde{K}^{-2\alpha_h})$ because

$$\mathbb{E}[\|\epsilon_{3,n}\||Z] \le C\sum_{1\le i,j\le n} \omega_{ij}\mathbb{E}[\tilde{\varepsilon}_i^2|Z]\left\|\tilde{h}_i\right\|^2/n \le C\sum_{1\le i,j\le n} \left\|\tilde{h}_i\right\|^2/n = O_{as}(\tilde{K}^{-2\alpha_h}).$$

Next, note that

$$\begin{aligned}\sum_{1\le i\le n}\|\tilde{h}_i\|^4 &\le \left(\sum_{1\le i\le n}\|\tilde{h}_i\|^2\right)^2 = O_{as}(n^2\tilde{K}^{-4\alpha_h}),\\ \frac{1}{n}\sum_{1\le i,j\le n}\omega_{ij}\tilde{g}_i^2\mathbb{E}[\|\tilde{v}_j\|^2|Z] &\le C\sum_{1\le i\le n}\tilde{g}_i^2/n = O_{as}(\tilde{K}^{-2\alpha_g}),\\ \frac{1}{n}\sum_{1\le i,j\le n}\tilde{g}_i^2\|\tilde{h}_j\|^2 &= \sum_{1\le i\le n}\tilde{g}_i^2\sum_{1\le j\le n}\|\tilde{h}_i\|^2/n = O_p(n\tilde{K}^{-2\alpha_h}\tilde{K}^{-2\alpha_g}).\end{aligned}$$

Therefore, by $|ab|\le Ca^2+Cb^2$, $\tilde{x}_i-\tilde{v}_i=\tilde{h}_i$, $\hat{\beta}-\beta=O_p\left(n^{-1/2}\right)$, and $\mathbb{E}[\|\tilde{v}_i\|^4|Z]\le C$,

$$\begin{aligned}\|\epsilon_{4,n}\| &\le C\sum_{1\le i,j\le n}\omega_{ij}\|\hat{\beta}-\beta\|^2\|\tilde{x}_i\|^2\|\tilde{x}_j\|^2/n + C\sum_{1\le i,j\le n}\omega_{ij}|\tilde{g}_i|^2\|\tilde{x}_j\|^2/n\\ &\le C\|\hat{\beta}-\beta\|^2\sum_{1\le i\le n}\|\tilde{v}_i\|^4/n + C\|\hat{\beta}-\beta\|^2\sum_{1\le i\le n}\|\tilde{h}_i\|^4/n\\ &\quad +C\sum_{1\le i,j\le n}\omega_{ij}\tilde{g}_i^2\|\tilde{v}_j\|^2/n + C\sum_{1\le i,j\le n}\tilde{g}_i^2\|\tilde{h}_i\|^2/n\\ &= O_p(n^{-1}+\tilde{K}^{-4\alpha_h}+\tilde{K}^{-2\alpha_g}+n\tilde{K}^{-2\alpha_g}\tilde{K}^{-2\alpha_h}) = o_p(1),\end{aligned}$$

Finally, $\epsilon_{5,n}=o_p(1)$ by the Cauchy-Schwarz inequality. Therefore, $\hat{\Upsilon}_n=\tilde{\Upsilon}_n+o_p(1)$.

Next, note that $\tilde{\Upsilon}_n=\vartheta_{1,n}+\vartheta_{2,n}+\vartheta_{3,n}$, where

$$\vartheta_{1,n}=\frac{1}{n}\sum_{1\le i,j\le n}\omega_{ij}\left(\sum_{1\le\ell\le n}\tilde{M}_{i\ell}^2\varepsilon_\ell^2\right)\left(\sum_{1\le\ell_1\le n}\tilde{M}_{j\ell_1}^2v_{\ell_1}v_{\ell_1}'\right),$$

$$\vartheta_{2,n}=\frac{1}{n}\sum_{1\le i,j\le n}\omega_{ij}\left(\sum_{1\le\ell\le n}\tilde{M}_{i\ell}^2\varepsilon_\ell^2\right)\left(\sum_{1\le\ell_1,\ell_2\le n,\ell_2\ne\ell_1}\tilde{M}_{j\ell_1}\tilde{M}_{j\ell_2}v_{\ell_1}v_{\ell_2}'\right),$$

$$\vartheta_{3,n}=\frac{1}{n}\sum_{1\le i,j\le n}\omega_{ij}\left(\sum_{1\le\ell_3,\ell_4\le n,\ell_3\ne\ell_4}\tilde{M}_{i\ell_3}\tilde{M}_{i\ell_4}\varepsilon_{\ell_3}\varepsilon_{\ell_4}\right)\left(\sum_{1\le\ell_1,\ell_2\le n}\tilde{M}_{j\ell_1}\tilde{M}_{j\ell_2}v_{\ell_1}v_{\ell_2}'\right),$$

with $\mathbb{E}\left[\vartheta_{2,n}|Z\right]=0=\mathbb{E}\left[\vartheta_{3,n}|Z\right]$. First, we show that $\vartheta_{1,n}=\mathbb{E}[\vartheta_{1,n}|Z]+o_p(1)$ with $\mathbb{E}[\vartheta_{1,n}|Z]=\bar{\Upsilon}_n$. Specifically, set $a_{1,ij}=\sum_{1\le\ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_2 j}^2$ to save notation, and note after expanding the sums, using the mean-zero property and collecting terms,

$$\begin{aligned}\mathbb{V}[\vartheta_{1,n}|Z] &= \frac{1}{n^2}\mathbb{E}\left[\left\|\sum_{1\le i,j\le n} a_{1,ij}\left(\varepsilon_i^2 v_j v_j' - \mathbb{E}[\varepsilon_i^2 v_j v_j'|Z]\right)\right\|^2 \middle| Z\right] \\ &\le \frac{C}{n^2}\sum_{1\le i,j,k\le n}[a_{1,ij}a_{1,ik} + a_{1,ik}a_{1,ji} + a_{1,ij}a_{1,jk} + a_{1,ik}a_{1,jk}] \le Cn^{-1},\end{aligned}$$

because, using properties of the idempotent matrices,

$$\begin{aligned}\sum_{1\le i,j,k\le n} a_{1,ij}a_{1,ik} &= \sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\sum_{i=1}^n \tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_3 i}^2\tilde{M}_{\ell_2\ell_2}\tilde{M}_{\ell_4\ell_4} \\ &\le C\sum_{1\le \ell_1\le n}\sum_{1\le \ell_3\le n}\sum_{i=1}^n \tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_3 i}^2 \le Cn,\end{aligned}$$

and similarly for the other terms. Thus, $\mathbb{V}[\|\vartheta_{1,n}\|^2|Z] \le Cn^{-1}$.

Next, $\vartheta_{2,n} = o_p(1)$ because $\mathbb{E}[\vartheta_{2,n}|Z] = 0$ and $\mathbb{E}[\|\vartheta_{2,n}\|^2|Z] \le Cn^{-1/2}$. In particular, let $a_{2,ijk} = \sum_{1\le\ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_2 j}\tilde{M}_{\ell_2 k}$, and observe that $a_{2,ijk} = a_{2,ikj}$. Expanding the sums, using the mean-zero properties of $\varepsilon_i^2 v_j v_k'$, and collecting terms gives $\mathbb{E}[\|\vartheta_{2,n}\|^2|Z] \le C\mathbb{E}[\|\vartheta_{21,n}\|^2|Z]/n^2 + C\mathbb{E}[\|\vartheta_{22,n}\|^2|Z]/n^2$, where

$$\mathbb{E}[\|\vartheta_{21,n}\|^2|Z] = \mathbb{E}\left[\left\|\sum_{1\le i,j\le n, j\ne i} a_{2,iji}\varepsilon_i^2 v_i v_j'\right\|^2 \middle| Z\right],$$

$$\mathbb{E}[\|\vartheta_{22,n}\|^2|Z] = \mathbb{E}\left[\left\|\sum_{1\le i,j,k\le n, j\ne i, k\ne i, k\ne j} a_{2,ijk}\varepsilon_i^2 v_j v_k'\right\|^2 \middle| Z\right].$$

For the first term ($\vartheta_{21,n}$), expanding the sums and collecting non-zero terms,

$$\begin{aligned}\mathbb{E}[\|\vartheta_{21,n}\|^2|Z]/n^2 &\le C\sum_{1\le i,j\le n, j\ne i}[a_{2,iji}^2 + a_{2,iji}a_{2,jij}]/n^2 + C\sum_{1\le i,j,k\le n, j\ne i, k\ne i, k\ne j} a_{2,iki}a_{2,jkj}/n^2 \\ &\le Cn^{-1} + Cn^{-1/2},\end{aligned}$$

where the second inequality uses $|a_{2,iji}a_{2,jij}| \le Ca_{2,iji}^2 + Ca_{2,jij}^2$ and

$$\begin{aligned}\sum_{1\le i,j\le n} a_{2,iji}^2 &\le \sum_{1\le i\le n}\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\left|\tilde{M}_{\ell_2 i}\right|\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 i}^2\left|\tilde{M}_{\ell_4 i}\right|\left|\sum_{1\le j\le n}\tilde{M}_{\ell_2 j}\tilde{M}_{\ell_4 j}\right| \\ &\le \sum_{1\le i\le n}\left(\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\left|\tilde{M}_{\ell_2 i}\right|\right)^2 \le C\sum_{1\le i\le n}\tilde{M}_{ii}^2 \le Cn\end{aligned}$$

for the first term, and also uses

$$\begin{aligned}&\sum_{1\le i,j,k\le n, j\ne i, k\ne i, k\ne j} a_{2,iki}a_{2,jkj} \\ &= \sum_{1\le i,j\le n}\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_2 i}\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 j}^2\tilde{M}_{\ell_4 j}\left(\sum_{1\le k\le n, k\ne i, k\ne j}\tilde{M}_{\ell_2 k}\tilde{M}_{\ell_4 k}\right),\end{aligned}$$

$$\sum_{k=1,k\ne i,k\ne j}^{n}\tilde{M}_{\ell_2 k}\tilde{M}_{\ell_4 k} = \tilde{M}_{\ell_2\ell_4} - \tilde{M}_{\ell_2 i}\tilde{M}_{\ell_4 i} - \tilde{M}_{\ell_2 j}\tilde{M}_{\ell_4 j},$$

$$\begin{aligned}&\sum_{1\le i,j\le n}\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\left|\tilde{M}_{\ell_2 i}\right|\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 j}^2\left|\tilde{M}_{\ell_4 j}\right|\left|\tilde{M}_{\ell_2\ell_4}\right| \\ &\le \sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\left|\tilde{M}_{\ell_2\ell_4}\right| \\ &\le C\sum_{1\le \ell_2,\ell_4\le n}\left|\tilde{M}_{\ell_2\ell_4}\right| \le C\sqrt{n}\sum_{1\le \ell_2\le n}\sqrt{\sum_{1\le \ell_4\le n}\tilde{M}_{\ell_2\ell_4}^2} \le Cn^{3/2},\end{aligned}$$

$$\begin{aligned}&\sum_{1\le i,j\le n}\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\left|\tilde{M}_{\ell_2 i}\right|\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 j}^2\left|\tilde{M}_{\ell_4 j}\right|\left|\tilde{M}_{\ell_2 i}\tilde{M}_{\ell_4 i}\right| \\ &\le \sum_{1\le i,j\le n}\sum_{1\le \ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\sum_{1\le \ell_3,\ell_4\le n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 j}^2\left|\tilde{M}_{\ell_4 i}\right| \le Cn^{3/2},\end{aligned}$$

for the second term.

Similarly, expanding the sums and collecting non-zero terms, we also obtain

$$\begin{aligned}\mathbb{E}[\|\vartheta_{22,n}\|^2 \,|Z]/n^2 &\leq C \sum_{1\leq i,j,k\leq n, j\neq i, k\neq i, k\neq j} [a_{2,ijk}^2 + a_{2,ijk}a_{2,jik}]/n^2 \\ &\quad + C \sum_{1\leq i,j,k,l\leq n, j\neq i,k\neq i,k\neq j,l\neq i,l\neq j,l\neq k} a_{2,ikl}a_{2,jkl}/n^2 \\ &\leq Cn^{-1} + Cn^{-1},\end{aligned}$$

where the second inequality uses $|a_{2,ijk}a_{2,jik}| \leq Ca_{2,ijk}^2 + a_{2,jik}^2$ and

$$\begin{aligned}\sum_{1\leq i,j,k\leq n} a_{2,ijk}^2 &= \sum_{1\leq i,j,k\leq n}\left(\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_2 j}\tilde{M}_{\ell_2 k}\right)^2 \\ &= \sum_{1\leq i\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2 \sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3 i}^2\tilde{M}_{\ell_2\ell_4}^2 \\ &\leq C\sum_{1\leq i\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}^2\tilde{M}_{ii}\tilde{M}_{\ell_2\ell_2} \leq C\sum_{1\leq i\leq n}\tilde{M}_{ii}\leq Cn\end{aligned}$$

for the first term, and also uses

$$\begin{aligned}&\sum_{\substack{1\leq i,j,k,l\leq n\\ j\neq i,k\neq i,k\neq j,l\neq i,l\neq j,l\neq k}} a_{2,ikl}a_{2,jkl} \\ &= \sum_{\substack{1\leq i,j\leq n\\ j\neq i}}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2 i}^2\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4 j}^2\sum_{\substack{1\leq k\leq n\\ k\neq i,k\neq j}}\tilde{M}_{\ell_1 k}\tilde{M}_{\ell_3 k}\sum_{\substack{1\leq l\leq n\\ l\neq i,l\neq j,l\neq k}}\tilde{M}_{\ell_1 l}\tilde{M}_{\ell_3 l},\end{aligned}$$

$$\begin{aligned}&\sum_{1\leq k\leq n,k\neq i,k\neq j}\tilde{M}_{\ell_1 k}\tilde{M}_{\ell_3 k}\sum_{1\leq l\leq n,l\neq i,l\neq j,l\neq k}\tilde{M}_{\ell_1 l}\tilde{M}_{\ell_3 l} \\ &= (\tilde{M}_{\ell_1\ell_3} - \tilde{M}_{\ell_1 i}\tilde{M}_{\ell_3 i} - \tilde{M}_{\ell_1 j}\tilde{M}_{\ell_3 j})^2 - \sum_{1\leq k\leq n,k\neq i,k\neq j}\tilde{M}_{\ell_1 k}^2\tilde{M}_{\ell_3 k}^2,\end{aligned}$$

$$\sum_{1\leq i,j\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2 i}^2\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4 j}^2\tilde{M}_{\ell_1\ell_3}^2$$
$$=\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2\ell_2}\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4\ell_4}\tilde{M}_{\ell_1\ell_3}^2\leq C\sum_{1\leq \ell_1,\ell_3\leq n}\tilde{M}_{\ell_1\ell_3}^2\leq Cn,$$

$$\sum_{1\leq i,j\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2 i}^2\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4 j}^2\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_3 i}^2$$
$$=\sum_{1\leq i\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2 i}^2\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4\ell_4}\tilde{M}_{\ell_1 i}^2\tilde{M}_{\ell_3 i}^2$$
$$\leq C\sum_{1\leq i\leq n}\tilde{M}_{ii}\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4\ell_4}\tilde{M}_{\ell_3 i}^2\leq C\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_3\ell_3}\leq Cn,$$

$$\sum_{1\leq i,j\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2 i}^2\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4 j}^2\sum_{1\leq k\leq n}\tilde{M}_{\ell_1 k}^2\tilde{M}_{\ell_3 k}^2$$
$$\leq C\sum_{1\leq k\leq n}\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_2\ell_2}\sum_{1\leq \ell_3,\ell_4\leq n}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_4\ell_4}\tilde{M}_{\ell_1 k}^2\tilde{M}_{\ell_3 k}^2$$
$$\leq C\sum_{1\leq k\leq n}\sum_{1\leq \ell_1,\ell_3\leq n}\tilde{M}_{\ell_1 k}^2\tilde{M}_{\ell_3 k}^2=\sum_{1\leq k\leq n}\tilde{M}_{kk}^2\leq Cn,$$

for the second term. Therefore, $\mathbb{E}[\|\vartheta_{2,n}\|^2|Z]\leq Cn^{-1/2}$, which implies $\vartheta_{2,n}=o_p(1)$.

Finally, $\vartheta_{3,n}=o_p(1)$ because $\mathbb{E}[\vartheta_{3,n}|Z]=0$ and $\mathbb{E}[\|\vartheta_{3,n}\|^2|Z]\leq Cn^{-1}$. To see the last conclusion, first let

$$a_{3,ij}=\sum_{1\leq \ell_1,\ell_2\leq n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}\tilde{M}_{\ell_1 j}\left(\sum_{1\leq k_1,k_2\leq n}\tilde{M}_{\ell_2 k_1}\tilde{M}_{\ell_2 k_2}v_{k_1}v_{k_2}'\right),$$

and observe that expanding the sums and collecting non-zero terms,

$$\begin{aligned}\mathbb{E}[\|\vartheta_{3,n}\|^2\,|Z] &= \mathbb{E}\left[\left\|\sum_{1\le i,j\le n,i\ne j} a_{3,ij}\varepsilon_i\varepsilon_j/n\right\|^2\middle| Z\right] = C\sum_{1\le i,j\le n,i\ne j}\mathbb{E}[\|a_{3,ij}\varepsilon_i^2\varepsilon_j^2\|^2|Z]/n^2\\ &\le \frac{C}{n^2}\sum_{1\le i,j\le n,j\ne i}\mathbb{E}\left[\left\|\sum_{1\le k_1,k_2,\ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}\tilde{M}_{\ell_1 j}\tilde{M}_{\ell_2 k_1}\tilde{M}_{\ell_2 k_2}v_{k_1}v_{k_2}'\right\|^2\middle| Z\right].\end{aligned}$$

Next, for each $(i,j)$ let

$$b_{3,ij,kl} = \sum_{1\le\ell_1,\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1 i}\tilde{M}_{\ell_1 j}\tilde{M}_{\ell_2 k}\tilde{M}_{\ell_2 l},$$

and note that $b_{3,ij,kl} = b_{3,ij,lk}$. Thus, expanding the sums and collecting non-mean-zero terms,

$$\begin{aligned}\mathbb{E}[\|\vartheta_{3,n}\|^2\,|Z] &\le \frac{C}{n^2}\sum_{1\le i,j\le n,j\ne i}\mathbb{E}\left[\left\|\sum_{1\le k\le n} b_{3,ij,kk}v_k v_k'\right\|^2\middle| Z\right]\\ &\quad+\frac{C}{n^2}\sum_{1\le i,j\le n,j\ne i}\mathbb{E}\left[\left\|\sum_{1\le k,l\le n,l\ne k} b_{3,ij,kl}v_k v_l'\right\|^2\middle| Z\right]\\ &\le \frac{C}{n^2}\sum_{1\le i,j,k\le n,j\ne i} b_{3,ij,kk}^2 + \frac{C}{n^2}\sum_{1\le i,j,k,l\le n,j\ne i,l\ne k} b_{3,ij,kl}^2 \le Cn^{-1},\end{aligned}$$

because

$$\begin{aligned}\sum_{1\le i,j,k,l\le n} b_{3,ij,kl}^2 &= \sum_{1\le\ell_1,\ell_2,\ell_3,\ell_4\le n}\omega_{\ell_1\ell_2}\omega_{\ell_3\ell_4}\tilde{M}_{\ell_1\ell_3}^2\tilde{M}_{\ell_2\ell_4}^2\\ &\le C\sum_{1\le\ell_1\le n}\sum_{1\le\ell_2\le n}\omega_{\ell_1\ell_2}\tilde{M}_{\ell_1\ell_1}\tilde{M}_{\ell_2\ell_2}\le Cn.\end{aligned}$$

This concludes the proof. Q.E.D.

**Proof of Theorem 3**. Note that $\tilde{M}_{i\ell_1}^2\tilde{M}_{j\ell_2}^2 = (\mathbf{1}(i=\ell_1)-\tilde{Q}_{i\ell_1})^2(\mathbf{1}(j=\ell_2)-\tilde{Q}_{j\ell_2})^2 = \mathbf{1}(i=\ell_1)\,\mathbf{1}(j=\ell_2)(1-2\tilde{Q}_{ii})(1-2\tilde{Q}_{jj})+\mathbf{1}(i=\ell_1)(1-2\tilde{Q}_{ii})\tilde{Q}_{j\ell_2}^2+\mathbf{1}(j=\ell_2)(1-2\tilde{Q}_{jj})\tilde{Q}_{i\ell_1}^2+\tilde{Q}_{i\ell_1}^2\tilde{Q}_{j\ell_2}^2$.

Next, $\mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j]$ is bounded, so that by $\tilde{Q}_{ii} \le 1$ and $M_{ij} = M_{ji}$,

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{1 \le i,j \le n} M_{ij}^2 \{(1 - 2\tilde{Q}_{ii})(1 - 2\tilde{Q}_{jj}) - 1\} \mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j] \right\| \\ \le \; & \frac{C}{n} \sum_{1 \le i,j \le n} M_{ij}^2 (\tilde{Q}_{ii} + \tilde{Q}_{jj}) \le \frac{C}{n} \sum_{1 \le j \le n} M_{jj} \tilde{Q}_{jj} \le \frac{C}{n} \sum_{1 \le j \le n} \tilde{Q}_{jj} \le C\tilde{K}/n \to 0, \end{aligned}$$

and, similarly, we also have

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{1 \le i,j \le n} \left( \sum_{1 \le \ell_1, \ell_2 \le n} M_{\ell_1 \ell_2}^2 \mathbf{1}(i = \ell_1)(1 - 2\tilde{Q}_{ii}) \tilde{Q}_{j\ell_2}^2 \right) \mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j] \right\| \\ \le \; & \frac{C}{n} \sum_{1 \le i,j,\ell \le n} M_{i\ell}^2 \tilde{Q}_{j\ell}^2 = \frac{C}{n} \sum_{1 \le \ell \le n} M_{\ell\ell} \tilde{Q}_{\ell\ell} \to 0, \end{aligned}$$

$$\begin{aligned} & \left\| \frac{1}{n} \sum_{1 \le i,j \le n} \left( \sum_{1 \le \ell_1, \ell_2 \le n} M_{\ell_1 \ell_2}^2 \tilde{Q}_{i\ell_1}^2 \tilde{Q}_{j\ell_2}^2 \right) \mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j] \right\| \\ \le \; & \frac{C}{n} \sum_{1 \le i,j \le n} \sum_{1 \le \ell_1, \ell_2 \le n} M_{\ell_1 \ell_2}^2 \tilde{Q}_{i\ell_1}^2 \tilde{Q}_{j\ell_2}^2 = \frac{C}{n} \sum_{1 \le \ell_1, \ell_2 \le n} M_{\ell_1 \ell_2}^2 \tilde{Q}_{\ell_1 \ell_1} \tilde{Q}_{\ell_2 \ell_2} \le \frac{C}{n} \left( \sum_{1 \le \ell_1, \ell_2 \le n} M_{\ell_1 \ell_2}^2 \tilde{Q}_{\ell_2 \ell_2} \right) \\ = \; & \frac{C}{n} \sum_{1 \le \ell_2 \le n} M_{\ell_2 \ell_2} \tilde{Q}_{\ell_2 \ell_2} \to 0. \end{aligned}$$

Therefore it follows that $\bar{\Upsilon}_n = \Sigma_n + o_p(1)$, so the conclusion follows by Lemma 4. Q.E.D.

**Proof of Theorem 4**. Follows by similar arguments to those given in the proof of Theorem 3. Q.E.D.

**Proof of Theorem 5**: Note first that when $\tilde{K} = K$ and $\omega_{i,k} = \omega_{ij} 1(i = j)$ we have $\hat{\Upsilon} = \hat{\Sigma}_k$. Then by the conclusion of Lemma 4

$$\hat{\Sigma}_k = \bar{\Upsilon}_n + o_p(1), \qquad \bar{\Upsilon}_n = \frac{1}{n} \sum_{1 \le i,j \le n} \left( \sum_{1 \le \ell \le n} \omega_{\ell,k} M_{i\ell}^2 M_{j\ell}^2 \right) \mathbb{E}[v_i v_i' \varepsilon_j^2 | z_i, z_j].$$

Under homoskedasticity we have

$$\begin{aligned}
\bar{\Upsilon}_n &= \frac{\sigma_\varepsilon^2}{n} \sum_{1\le i,j\le n} \left( \sum_{1\le \ell\le n} \omega_{\ell,k} M_{i\ell}^2 M_{j\ell}^2 \right) \mathbb{E}[v_i v_i' | z_i] = \frac{\sigma_\varepsilon^2}{n} \sum_{1\le i,j\le n} \omega_{j,k} M_{jj} M_{ji}^2 \mathbb{E}[v_i v_i' | z_i], \\
\Sigma_n &= \frac{\sigma_\varepsilon^2}{n} \sum_{1\le i,j\le n} M_{ij}^2 \mathbb{E}[v_i v_i' | z_i].
\end{aligned}$$

Substituting $\omega_{j,k} = 1$ and subtracting gives the first result. Substituting $\omega_{j,k} = M_{jj}^{-1}$ gives the second conclusion for the case $\delta = 1$. Next, note that $M_{ii}^{-1} > 1$, and hence

$$M_{ji}^2 \mathbb{E}[v_i v_i' | z_i] < M_{jj}^{-1} M_{ji}^2 \mathbb{E}[v_i v_i' | z_i] = M_{jj}^{-\delta+1} M_{jj} M_{ji}^2 \mathbb{E}[v_i v_i' | z_i], \qquad \delta > 1.$$

Summing up gives the conclusion. Q.E.D.

# Research Papers 2012

2011-40: Søren Johansen and Bent Nielsen: Asymptotic theory for iterated one-step Huber-skip estimators

2011-41: Luc Bauwens, Arnaud Dufays and Jeroen V.K. Rombouts: Marginal Likelihood for Markov-switching and Change-point Garch Models

2011-42: Manuel Lukas: Utility-based Forecast Evaluation with Multiple Decision Rules and a New Maxmin Rule

2011-43: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, Mehdi Karoui: Illiquidity Premia in the Equity Options Market

2011-44: Diego Amaya, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez: Do Realized Skewness and Kurtosis Predict the Cross-Section of Equity Returns?

2011-45: Peter Christoffersen and Hugues Langlois: The Joint Dynamics of Equity Market Factors

2011-46: Peter Christoffersen, Kris Jacobs and Bo Young Chang: Forecasting with Option Implied Information

2011-47: Kim Christensen and Mark Podolskij: Asymptotic theory of range-based multipower variation

2011-48: Christian M. Dahl, Daniel le Maire and Jakob R. Munch: Wage Dispersion and Decentralization of Wage Bargaining

2011-49: Torben G. Andersen, Oleg Bondarenko and Maria T. Gonzalez-Perez: Coherent Model-Free Implied Volatility: A Corridor Fix for High-Frequency VIX

2011-50: Torben G. Andersen and Oleg Bondarenko: VPIN and the Flash Crash

2011-51: Tim Bollerslev, Daniela Osterrieder, Natalia Sizova and George Tauchen: Risk and Return: Long-Run Relationships, Fractional Cointegration, and Return Predictability

2011-52: Lars Stentoft: What we can learn from pricing 139,879 Individual Stock Options

2011-53: Kim Christensen, Mark Podolskij and Mathias Vetter: On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes

2012-01: Matei Demetrescu and Robinson Kruse: The Power of Unit Root Tests Against Nonlinear Local Alternatives

2012-02: Matias D. Cattaneo, Michael Jansson and Whitney K. Newey: Alternative Asymptotics and the Partially Linear Model with Many Regressors