Bootstrap Inference for Partially Linear Model with Many Regressors

Wang, Wenjie

3 March 2021

Bootstrap Inference for Partially Linear Model with Many Regressors

Wenjie Wang *

March 3, 2021

Abstract

In this note, for the case that the disturbances are conditional homoskedastic, we show that a properly re-scaled residual bootstrap procedure is able to consistently estimate the limiting distribution of a series estimator in the partially linear model even when the number of regressors is of the same order as the sample size. Monte Carlo simulations show that the bootstrap procedure has superior finite sample performance than asymptotic approximations when the sample size is small and the number of regressors is close to the sample size.

Keywords: Bootstrap approximation, Partially linear model, Many regressors asymptotics

*Division of Economics, School of Social Sciences, Nanyang Technological University, HSS-04-65, 14 Nanyang Drive, 637332, Singapore. E-mail address: wang.wj@ntu.edu.sg.

1 Introduction

In recent years, various efforts have been made to relax linear regression model assumptions and hence widen their applicability, since a wrong model on the regression function can lead to excessive modeling biases and erroneous conclusion. Of importance is the partially linear regression model which allows the relationship between response and partial covariates to be not specified. A partially linear regression model can be written as

$$y_i = x_i\beta + g(z_i) + \epsilon_i, i = 1, ..., n$$

where β is parameter of interest and $g(\cdot)$ is unknown function. Donald and Newey (1994) discussed estimating the model using series approximation (e.g., polynomials or splines) to $g(z_i)$ and gave conditions for the asymptotic normality of this estimator using standard asymptotics. However, this result is not robust to the number of terms in the series approximation since it is assumed to be fixed when the number of sample size goes to infinity. Recently, Cattaneo, Jansson and Newey (2018) propose an alternative asymptotic framework which allows the number of terms to grow as fast as the sample size. They show that under many regressors asymptotics, the limiting distribution of the series estimator has a larger than usual asymptotic variance. Moreover, when the disturbance is homoskedastic, this larger variance can be consistently estimated by the usual variance estimator provided that a degrees-of-freedom correction is used.

However, as can be seen from our simulation result, inference based on the asymptotic normal approximation of Cattaneo et al. (2018) can have size distortion in finite samples, especially when the sample size is small and the number of regressors is close to the sample size. It is therefore tempting to consider whether other methods, such as the bootstrap, is able to provide a better alternative than their asymptotic normal approximation. Indeed, in a similar context of linear instrumental variable model where the number of instruments is allowed to be a nontrivial fraction of sample size, the bootstrap, when designed properly, is found to be asymptotically valid and has better finite-sample performance than conventional asymptotic approximations in terms of size control for hypothesis testing; e.g., see Wang and Kaffo (2016), Kaffo and Wang (2017), and Wang (2020).

In this paper, we propose a residual-based i.i.d. bootstrap which puts mass $1/n$ at each (re-scaled) residual in order to approximate the limiting distribution of series estimator. The residuals are properly re-scaled to account for the affect of using a large number of regressors. We show analytically that our bootstrap techniques provide a valid method to approximate the limiting distribution of the series estimator under Cattaneo et al. (2018)'s many regressors asymptotics. A Monte Carlo experiment shows that confidence intervals based on Cattaneo et al. (2018)'s normal approximation can have size distortion in finite sample, especially when the number of regressors is large. Our bootstrap procedure reduces these distortions.

The rest of this paper is organized as follows. Section 2 describes setup of the model and the many regressor asymptotics in the partially linear model. Then, we present the result for bootstrap-based inference on the partially linear model with many series regressors. In Section 3, we present some Monte Carlo simulation results, while Section 4 concludes.

2 Setup and Main Results

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

$$y_i = x_i\beta + g(z_i) + \epsilon_i, E[\epsilon_i|x_i, z_i] = 0,$$

and $x_i = h(z_i) + v_i$ with $h(z_i) = E[x_i|z_i]$; $g(\cdot)$ and $h(\cdot)$ are unknown functions. We also assume that the disturbances are homoskedastic: $E[\epsilon_i^2|x_i, z_i] = \sigma_{\epsilon\epsilon},\ E[v_i^2|z_i] = \Sigma_{vv}$. Following Cattaneo et al. (2018), we will condition on Z throughout the following discussion (alternatively, we could assume that Z is non-random, as pointed out by Cattaneo et al. (2018).

A series estimator of β is obtained by using approximating function of z_i. To describe the estimator, let $p^k(z) = (p_{1k}(z), ..., p_{kk}(z))'$ 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 the approximating functions. Therefore, letting $Y = [y_1, ..., y_n]' \in R^{n\times 1}$ and $X = [x_1, ..., x_n]' \in R^{n\times 1}$, and $P_Z = [p^k(z_1), ..., p^k(z_n)]'$, a series estimator of β is given by

$$\hat{\beta} = (X'M_ZX)^{-1}X'M_ZY, M_Z = I - Q_Z, Q_Z = P_Z(P_Z'P_Z)^{-}P_Z'$$

where A^- denotes a generalized inverse of a matrix A (satisfying $AA^-A = A$) and $X'M_ZX$ will be non-singular with probability approaching one under appropriate conditions.

The limiting distribution of $\hat{\beta}$ was derived by Donald and Newey (1994) under conventional asymptotics where the number of series terms is assumed to be fixed. However, such asymptotics cannot provide a good approximation for the distribution of $\hat{\beta}$ when the number of regressors become large.

To obtain a better approximation in the case of many regressors, Cattaneo et al. (2018) recently proposed an alternative asymptotic framework, in which the number of terms in the series approximation is allowed to grow as fast as the sample size. More precisely, it is shown that when the disturbances are homoskedastic,

$$\sqrt{n}(\hat{\beta} - \beta) \to^d N(0, \Omega)$$

under many regressors asymptotics, where $\Omega = \sigma_{\epsilon\epsilon}\Gamma^{-1}, \Gamma = (1-\alpha)\Sigma_{vv}$ with $\alpha \in [0, 1)$. Apparently, the presence of many regressors leads to a larger asymptotic variance which captures term assumed away by the classical asymptotic result. Moreover, Cattaneo et al. (2018) show that this larger variance can be consistently estimated by using the usual asymptotic variance estimator with a proper degrees of freedom correction. In particular, they show that

$$\widehat{\Omega}^{-1/2}(\hat{\beta}-\beta) \to^d N(0,1)$$

where $\widehat{\Omega} = \hat{s}^2\widehat{\Gamma}^{-1}, \hat{s}^2 = \hat{\epsilon}'\hat{\epsilon}/(n-k-1), \hat{\epsilon} = M_Z(Y - X\hat{\beta}), \widehat{\Gamma} = X'M_Z X/n.$

However, as can be seen from our simulation result, inference based on Cattaneo et al. (2018)'s asymptotic approximation can still have serious distortion, especially when the number of regressors is large relative to the sample size. Here, we propose a bootstrap method as an alternative to their asymptotic normal approximation. We show both analytically and numerically that this method is valid under many regressors asymptotics in the sense that our method consistently estimate the limiting distribution of $\hat{\beta}$ even when the number of terms in the series approximation possibly grows as fast as the sample size.

Specifically, our bootstrap procedure proceeds as follows.

1. We obtain properly re-scaled residuals

$$\hat{\epsilon} = M_Z(Y - X\hat{\beta}),\ \hat{V} = M_Z X$$

   Denote $\bar{\epsilon} = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$. Let $\hat{F}_n^{\epsilon}$ be the empirical distribution of $\sqrt{\frac{n}{n-k}}\,(\hat{\epsilon}_i - \bar{\epsilon})$, so $\hat{F}_n^{\epsilon}$ puts mass $\frac{1}{n}$ at $\sqrt{\frac{n}{n-k}}\,(\hat{\epsilon}_i - \bar{\epsilon})$ and $\int x d\hat{F}_n^{\epsilon}(x) = 0$. Similarly, denote $\bar{V} = n^{-1}\sum_{i=1}^n \hat{V}_i$. Let $\hat{F}_n^v$ be the empirical distribution of $\sqrt{\frac{n}{n-k}}\left(\hat{V}_i - \bar{V}\right)$, so $\hat{F}_n^v$ puts mass $\frac{1}{n}$ at $\sqrt{\frac{n}{n-k}}\left(\hat{V}_i - \bar{V}\right)$ and $\int x d\hat{F}_n^v(x) = 0$.

2. Generate $\{\epsilon_1^*, ..., \epsilon_n^*\}$, which are conditionally independent with common distribution $\hat{F}_n^{\epsilon}$ and $\{V_1^*, ..., V_n^*\}$, which are conditionally independent with common distribution $\hat{F}_n^v$.

3. We set the bootstrap data generating process (DGP) as

$$\begin{aligned} X^* &= Q_Z X + V^* \\ Y^* &= X^*\hat{\beta} + Q_Z(y - X\hat{\beta}) + \epsilon^* \end{aligned}$$

   where $\epsilon^* = (\epsilon_1^*, ..., \epsilon_n^*)'$ and $V^* = (V_1^*, ..., V_n^*)'$.

4. Generate $\hat{\beta}^* = \left(X^{*\prime} M_Z X^*\right)^{-1} X^{*\prime} M_Z Y^*$ using bootstrap pseudo-data.

Remark 1. Note that in our bootstrap procedure, the residuals $(\hat{\epsilon}, \hat{v})$ are properly re-scaled to take account for the affect of using many terms in the series approximation. In the bootstrap literature, such degree-of-freedom correction are often implemented to obtain a better finite sample performance. Here, our result gives a large sample justification for the correction because under many regressors asymptotics, this is no longer a problem of finite sample performance, and re-scaling the residuals becomes essential to establishing the validity of residual-based bootstrap, as will be shown in Theorem 2.1 below. 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.

In order to formally investigate the asymptotic properties of the bootstrap estimator $\hat{\beta}^*$, we begin with the following assumptions. These assumptions are also used in Cattaneo et al. (2018).

Assumption 1.

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

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

As pointed out by Cattaneo et al. (2018), these conditions are implied by conventional assumptions from approximation theory. They are needed for the control of bias from approximating unknown functions by a linear combination of $p^k(\cdot)$. Next, we also assume that certain moments of the disturbances are bounded.

Assumption 2.

There is $C < \infty$ such that $E[V_i^4] \leq C$ and $E[\epsilon_i^4] \leq C$.

We are now ready to establish the main results.

Theorem 2.1 *Suppose that Assumptions 1-2 hold. Then, if $\alpha_n = k/n \to \alpha \in [0,1)$,*

$$sup_{x\in R}\left|P^*\left[\sqrt{n}(\hat{\beta}^* - \hat{\beta}) \leq x\right] - P\left[\sqrt{n}(\hat{\beta} - \beta) \leq x\right]\right| \to^p 0,$$

and

$$sup_{x\in R}\left|P^*\left[\widehat{\Omega}^{*-1/2}(\hat{\beta}^* - \hat{\beta}) \leq x\right] - P\left[\widehat{\Omega}^{-1/2}(\hat{\beta} - \beta) \leq x\right]\right| \to^p 0$$

where P^* denotes the probability measure induced by the bootstrap procedure proposed in this section.

Remark 2. The first result in Theorem 2.1 guarantees the asymptotic validity of percentile type tests and confidence intervals (CIs) based on our bootstrap method. More precisely, Percentile type bootstrap intervals based on $\hat{\beta}$ and $\hat{\beta}^*$ are computed as

$$\hat{\beta} \pm q^*_{0.95}, \tag{1}$$

where $q^*_{0.95}$ is such that $P^*\left(\left|\hat{\beta}^* - \hat{\beta}\right| \leq q^*_{0.95}\right) = 0.95$. Moreover, the second result in Theorem 2.1 shows that percentile-t tests and CIs based on our bootstrap method is also valid. Therefore, we can construct CI for $\hat{\beta}$ using

$$\hat{\beta} \pm z^*_{0.95}\sqrt{\widehat{\Omega}}, \tag{2}$$

where $z^*_{0.95}$ is such that $P^*\left(\left|\frac{\hat{\beta}^* - \hat{\beta}}{\sqrt{\widehat{\Omega}^*}}\right| \leq z^*_{0.95}\right) = 0.95$. And we can define percentile and percentile-t type bootstrap tests accordingly.

Remark 3. It is easy to see that our bootstrap is also valid under Donald and Newey (1994)'s conventional asymptotics where $\alpha_n = k/n \to 0$. Therefore, this procedure can be seen as a unified inference approach which is valid regardless of the number of regressors used in the series estimation.

Table 1. Empirical rejection frequency at 5 percent nominal level.

	$n = 100$			$n = 200$		
	$k = 10$	$k = 40$	$k = 70$	$k = 10$	$k = 40$	$k = 70$
$t_{HO,1}$	10.7	6.7	1.4	10.6	8.3	6.6
$t_{HO,2}$	9.0	1.8	0	9.7	5.2	2.1
$\hat{\beta}^*_{part}$	5.3	5.0	5.8	5.8	4.7	5.7
$\hat{\beta}^*_{lhs}$	5.6	3.0	0.3	5.7	4.0	2.8

Note: $t_{HO,1}$ and $t_{HO,2}$ denote the t-ratios studied by Cattaneo et al. (2018). $\hat{\beta}^*_{part}$ denotes the i.i.d. bootstrap proposed in this note. $\hat{\beta}^*_{lhs}$ denotes the i.i.d. bootstrap proposed in Liang et al. (2000).

3 Simulation

We conduct a Monte Carlo experiment to explore the finite sample performance of the bootstrap procedures proposed in the previous sections. Throughout this section, the simulation study is based on 5000 replications. We set sample size $n = 100$ and $n = 200$ and we set the number of bootstrap replication as $B = 399$.

For the partially linear model, we consider the following setting:

$$\begin{aligned} y_i &= x_i\beta + g(z_i) + \epsilon_i \\ x_i &= h(z_i) + v_i \end{aligned}$$

where $\beta = 0, z_i \sim U(-1, 1), \epsilon_i \sim N(0, 1), v_i \sim U(-1, 1)$, and $g(z_i) = z_i(2 + z_i)^{-1/2}$.

The estimator considered in the Monte Carlo experiment is based on power series approximation. Specifically, we approximate $g(z_i)$ by $p^k(z_i)'\gamma$ with $p^k(z_i) = (1, z_i, z_i^2, ..., z_i^k)'$. For the choices of k, the number of regressors in the partially linear model, we set $k = 10, 40, 70$. To explore the consequences of introducing many regressors in the partially linear model, we focus on the finite sample size properties of 4 competing methods: $t_{HO,m}$, m=1,2, which denote the two t-ratios studies in Cattaneo et al. (2018), the percentile type bootstrap test based on the residual i.i.d. procedure proposed in this note, and the percentile type bootstrap test based on the procedure proposed in Liang et al. (2000).

The main finding from the simulation is presented in Table 1. It turns out that the finite sample rejection frequency of $t_{HO,m}, m = 1, 2$ is quite sensitive to the choice of k. Specifically, these tests tend to over-reject the null hypothesis when k is relatively small, but tend to under-reject when k becomes large. Also, the performance of $t_{HO,1}$ is not always superior to that of $t_{HO,2}$, when k is large, the degrees of freedom corrected $t_{HO,1}$ has difficulty to reject the null hypothesis. For the two bootstrap procedures, Liang, Härdle, and Sommerfeld (2000)'s bootstrap is able to well control the size when k is small, but it also under-rejects when k becomes a non-trivial fraction of the sample size. In contrast, Our bootstrap procedure has close-to-correct empirical size for the full range of k in the simulation.

4 Conclusion

In this note, we employ the asymptotic result introduced by Cattaneo et al. (2018) to derive results concerning bootstrap based inference for the partially linear model and the instrumental variable model. For the homoskedastic case, We show that when the residuals are properly re-scaled to account for the presence of many regressors, the i.i.d. bootstrap is able to well mimic the limiting distribution of the series estimator in the partially linear model, even when the number of regressors goes to infinity at the same speed as the sample size. For future research, we shall investigate the higher order properties of the bootstrap procedure when many regressors are involved by investigating the property of edgeworth expansions in such circumstances.

References

CATTANEO, M. D., M. JANSSON, AND W. K. NEWEY (2012): "Alternative Asymptotics and the Partially Linear Model with Many Regressors," CREATES Research Papers 2012-02, School of Economics and Management, University of Aarhus.

CATTANEO, M. D., M. JANSSON, AND W. K. NEWEY (2018): "Alternative asymptotics and the partially linear model with many regressors," *Econometric Theory*, 34(2), 277–301.

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. NEWEY (2008): "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422.

KAFFO, M., AND W. WANG (2017): "On bootstrap validity for specification testing with many weak instruments," *Economics Letters*, 157, 107–111.

LIANG, H., W. HÄRDLE, AND V. SOMMERFELD (2000): "Bootstrap approximation in a partially linear regression model," *Journal of statistical planning and inference*, 91(2), 413–426.

WANG, W. (2020): "On Bootstrap Validity for the Test of Overidentifying Restrictions with Many Instruments and Heteroskedasticity," Discussion Paper No. 104858, University Library of Munich, Germany.

WANG, W., AND M. KAFFO (2016): "Bootstrap inference for instrumental variable models with many weak instruments," *Journal of Econometrics*, 192(1), 231–268.

A Mathematical Proofs

Throughout this Appendix, let C be a generic positive constant that may be different in different use. For any bootstrap statistic T^* we write $T^* \to^{p^*} 0$ in probability when $lim_{n\to\infty} P[P^*(|T^*| > \delta) > \delta] = 0$ for any $\delta > 0$, i.e. $P^*(|T^*| > \delta) = o_p(1)$. Also, we say that $T^* = O_{p^*}(n^\lambda)$ in probability if and only if $\forall \delta > 0$, There exists a $M_\delta < \infty$ such that $lim_{n\to\infty} P[P^*(|n^{-\lambda} T^*| > M_\delta) > \delta] = 0$, i.e. $\forall \delta > 0$, There exists a $M_\delta < \infty$ such that $P^*(|n^{-\lambda} T^*| > M_\delta) = o_p(1)$. Finally, we write $T^* \to^{d^*} D$ in probability, for any distribution D, when weak convergence under the bootstrap probability measure occurs in a set with probability converging to one.

Lemma A.1 *If Assumption 1 is satisfied, then if for some $\delta > 0$, $E[V_i^4] < \infty$ and $E[\epsilon_i^4] < \infty$, then $E^*\left[V_i^{*4}\right]$ and $E^*\left[\epsilon_i^{*4}\right]$ are bounded in probability.*

Proof of Lemma A.1

We give the proof for $E^*\left[V_i^{*4}\right]$, the proof for $E^*\left[\epsilon_i^{*4}\right]$ is similar.

$$
\begin{aligned}
E^*\left[V_i^{*4}\right] &= \frac{1}{n}\sum_{i=1}^{n}\left[\sqrt{\frac{n}{n-k}}\left(\hat{V}_i - \bar{V}\right)\right]^4 \\
&= \left(\frac{1}{1-\alpha_n}\right)^4 \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{h}_i + \tilde{V}_i - \bar{V}\right)^4 \\
&\leq C\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{h}_i\right)^4 + \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{V}_i - \bar{V}\right)^4\right\}
\end{aligned}
$$

where $\alpha_n = k/n$, $\tilde{h}_i = h_i - \sum_{j=1}^{n} Q_{ij} h_j$ and $\tilde{V}_i = V_i - \sum_{j=1}^{n} Q_{ij} V_j$; Q_{ij} denotes the ijth element of Q_Z. The inequality follows from Minkowski's inequality.

Furthermore, note that

$$
\sum_{i=1}^{n}\tilde{h}_i^2 = trace\left(H' M_Z H\right) \leq \sum_{i=1}^{n}\left(h(z_i) - p^k(z_i)'\eta_h\right)^2 = O_{as}\left(nk^{-2\alpha_h}\right)
$$

by Assumption 1. Therefore, $\sum_{i=1}^{n}\tilde{h}_i^4 \leq \left(\sum_{i=1}^{n}\tilde{h}_i^2\right)^2 = O_{as}\left(n^2 k^{-4\alpha_h}\right)$. Moreover, we have

$$
E\left[\tilde{V}_i^4\right] \leq C\sum_{j=1}^{n}\sum_{l=1}^{n} M_{ij}^2 M_{il}^2 E\left[V_j^2 V_l^2\right] \leq C
$$

by properties of the idempotent matrix and by Assumption 2.

Therefore, we have by Minkowski's inequality

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{V}_i - \bar{V}\right)^4 &\leq C\left\{\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i^4 + \bar{V}^4\right\} \\
&\to^p C\left\{E\left[\tilde{V}_i^4\right]\right\}
\end{aligned}
$$

since $\bar{V}^4 \to^p (E[V_i])^4 = 0$. Putting these results together, we obtain

$$\begin{aligned} E^*\left[V_i^{*4}\right] &\leq C\left\{\frac{1}{n}\sum_{i=1}^n \tilde{h}_i^4 + \frac{1}{n}\sum_{i=1}^n \tilde{V}_i^4\right\} \\ &= O_{as}\left(nk^{-4\alpha_h}\right) + O_p(1) = O_p(1) \end{aligned}$$

by Assumption 1. i.e., $E^*[V_i^{*4}]$ is bounded in probability. ■

Lemma A.2 *If Assumption 1 and 2 are satisfied, then*

$$\hat{\Gamma}^* = \Gamma + o_{p^*}(1)$$

where $\hat{\Gamma}^* = X^{*'} M_Z X^*/n$*,* $\Gamma = (1-\alpha)\Sigma_{vv}$*.*

Proof of Lemma A.2

From the bootstrap DGP, we have

$$\begin{aligned} E^*\left[\frac{V^{*'}Q_Z V^*}{k}\right] &= \frac{1}{k}E^*\left[trace\left(V^{*'}Q_Z V^*\right)\right] = \frac{1}{k}trace\left(Q_Z E^*\left[V^* V^{*'}\right]\right) \\ &= \frac{trace(Q_Z)}{k}E^*\left[V_i^{*2}\right] = E^*\left[V_i^{*2}\right] \end{aligned}$$

since $E^*\left[V_i^* V_j^*\right] = E^*\left[V_i^*\right]E^*\left[V_j^*\right] = 0$ for $i \neq j$ by the property of i.i.d. bootstrap.

Furthermore, note that

$$\begin{aligned} & E^*\left[\frac{V^{*'}Q_Z V^*}{k} - E^*\left[V_i^{*2}\right]\right]^2 \\ = \;& \frac{1}{k^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\sum_{m=1}^n Q_{ij}Q_{lm}E^*\left[V_i^*V_j^*V_l^*V_m^*\right] - \frac{2}{k}E^*\left[V_i^{*2}\right]\sum_{i=1}^n\sum_{j=1}^n Q_{ij}E^*\left[V_i^*V_j^*\right] + \left(E^*\left[V_i^{*2}\right]\right)^2 \\ = \;& \frac{1}{k^2}E^*\left[V_i^{*4}\right]\sum_{i=1}^n (Q_{ii})^2 + \frac{2}{k^2}\left(E^*\left[V_i^{*2}\right]\right)^2\sum_{i=2}^n\sum_{j=1}^{i-1}(Q_{ij})^2 \\ & + \left\{\frac{2}{k^2}\left(E^*\left[V_i^{*2}\right]\right)^2\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\left(Q_{ii}Q_{jj} + (Q_{ij})^2\right)\right] - \left(E^*\left[V_i^{*2}\right]\right)^2\right\} \\ \equiv \;& L_1^* + L_2^* + L_3^* \end{aligned}$$

Let us first focus on L_1^*. Note that

$$L_1^* \leq \frac{1}{k}E^*\left[V_i^{*4}\right] = O_p\left(\frac{1}{k}\right)$$

by Lemma A.1 and by the fact that $\sum_{i=1}^n (Q_{ii})^2 \leq \sum_{i=1}^n Q_{ii} = k$.

Next, for L_2^*, we note that by our bootstrap DGP

$$E^*[V_i^{*2}] = \left(\frac{n}{n-k}\right)\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{V}_i - \bar{V}\right)^2\right) = \left(\frac{n}{n-k}\right)\left(\frac{\hat{V}'\hat{V}}{n} - \left(\bar{V}\right)^2\right) \to^p \Sigma_{vv}$$

since $\frac{\hat{V}'\hat{V}}{n} \to^p (1-\alpha)\Sigma_{vv}$ and $\bar{V} \to^p E[V_i] = 0$. Moreover, we have

$$\sum_{i=1}^{n}(Q_{ii})^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(Q_{ij})^2 = Tr(Q_Z'Q_Z) = Tr(Q_Z) = k$$

given that Q_Z is symmetric and idempotent. Therefore,

$$L_2^* \leq \frac{1}{k^2}\left(E^*[V_i^{*2}]\right)^2\left[\sum_{i=1}^{n}(Q_{ii})^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(Q_{ij})^2\right] = \frac{1}{k}\left(E^*[V_i^{*2}]\right)^2 = O_p\left(\frac{1}{k}\right)$$

For L_3^*, we note that

$$\begin{aligned}
|L_3^*| &= \left|\frac{1}{k^2}\left(E^*[V_i^{*2}]\right)^2\left[(trace(Q_Z))^2 + trace(Q_Z'Q_Z) - 2\sum_{i=1}^{n}(Q_{ii})^2\right] - \left(E^*[V_i^{*2}]\right)^2\right| \\
&= \left|\frac{1}{k^2}\left(E^*[V_i^{*2}]\right)^2\left(k^2 + k - 2\sum_{i=1}^{n}(Q_{ii})^2\right) - \left(E^*[V_i^{*2}]\right)^2\right| \\
&= \left|\frac{1}{k^2}\left(E^*[V_i^{*2}]\right)^2\left(k - 2\sum_{i=1}^{n}(Q_{ii})^2\right)\right| \leq \frac{1}{k}\left(E^*[V_i^{*2}]\right)^2 + \frac{2}{k^2}\left(E^*[V_i^{*2}]\right)^2\sum_{i=1}^{n}Q_{ii} = O_p\left(\frac{1}{k}\right)
\end{aligned}$$

Putting these results together, we obtain $E^*\left[\frac{V^{*\prime}Q_ZV^*}{k} - E^*[V_i^{*2}]\right]^2 = O_p\left(\frac{1}{k}\right)$ and $\frac{V^{*\prime}Q_ZV^*}{k} - E^*[V_i^{*2}] = O_{p^*}\left(\frac{1}{\sqrt{k}}\right)$ by Markov's inequality.

Using similar arguments, we can show that $\frac{V^{*\prime}V^*}{n} - E^*[V_i^{*2}] = O_{p^*}\left(\frac{1}{\sqrt{n}}\right)$.

Finally, we obtain

$$\begin{aligned}
&\frac{V^{*\prime}M_ZV^*}{n} - (1-\alpha_n)\Sigma_{vv} \\
= &\frac{V^{*\prime}V^*}{n} - \frac{V^{*\prime}Q_ZV^*}{n} - (1-\alpha_n)\Sigma_{vv} \\
= &\left(\frac{V^{*\prime}V^*}{n} - E^*[V_i^{*2}]\right) - \alpha_n\left(\frac{V^{*\prime}Q_ZV^*}{k} - E^*[V_i^{*2}]\right) + (1-\alpha_n)\left(E^*[V_i^{*2}] - \Sigma_{vv}\right) \\
= &O_{p^*}\left(\frac{1}{\sqrt{n}}\right) - \alpha_nO_{p^*}\left(\frac{1}{\sqrt{k}}\right) + (1-\alpha_n)o_p(1) = o_{p^*}(1)
\end{aligned}$$

and the result follows because $X^{*\prime}M_ZX^*/n = (Q_ZX + V^*)'M_Z(Q_ZX + V^*)/n = V^{*\prime}M_ZV^*/n$. ■

Proof of Theorem 2.1

We proceed by checking that $\frac{1}{\sqrt{n}}V^{*\prime}M_Z\epsilon^*$ satisfies the condition of Lemma A2 in Hansen, Hausman and Newey (2008), conditionally on the original sample with probability converging to 1.

Let $W_i^* = \frac{1}{\sqrt{n}} M_{ii} V_i^* \epsilon_i^*$. First, by our bootstrap DGP, $\{W_i^*, V_i^*, \epsilon_i^*\}, i = 1, ..., n$ are (conditionally) independent across i. Second, $E^*[\epsilon_i^*] = n^{-1} \sum_{i=1}^n \left[\sqrt{\frac{n}{n-k}} (\hat{\epsilon}_i - \bar{\epsilon})\right] = 0, E^*[V_i^*] = n^{-1} \sum_{i=1}^n \left[\sqrt{\frac{n}{n-k}} \left(\hat{V}_i - \bar{V}\right)\right] = 0$. Third, $E^*[\epsilon_i^{*4}]$ and $E^*[V_i^{*4}]$ are bounded in probability by Lemma A.1. Fourth,

$$\begin{aligned}\sum_{i=1}^n E^*[W_i^{*4}] &= n^{-2} \sum_{i=1}^n M_{ii}^4 E^*[V_i^{*4} \epsilon_i^{*4}] \\ &= n^{-2} \sum_{i=1}^n M_{ii}^4 E^*[V_i^{*4}] E^*[\epsilon_i^{*4}] \\ &= O_p(1) n^{-2} \sum_{i=1}^n M_{ii}^4 \leq O_p(1) n^{-2} \sum_{i=1}^n M_{ii} = O_p(1) \frac{n-k}{n^2} \to^p 0\end{aligned}$$

where the second equality follows from the property of the bootstrap DGP, the third equality from Lemma A.1, and the inequality from the fact that $M_{ii}^4 \leq M_{ii}$ and $\sum_{i=1}^n M_{ii} = \sum_{i=1}^n (1 - Q_{ii}) = n - k$.

Finally, we can show that

$$\begin{aligned}& \sum_{i=1}^n E^*[W_i^{*2}] + \left(1 - \frac{\sum_{i=1}^n Q_{ii}^2}{k}\right) \alpha_n \left[E^*[V_i^{*2}] E^*[\epsilon_i^{*2}] + (E^*[V_i^* \epsilon_i^*])^2\right] \\ = \quad & \frac{\sum_{i=1}^n M_{ii}^2}{n} E^*[V_i^{*2}] E^*[\epsilon_i^{*2}] + \left(1 - \frac{\sum_{i=1}^n Q_{ii}^2}{k}\right) \alpha_n E^*[V_i^{*2}] E^*[\epsilon_i^{*2}] \\ = \quad & \left[\frac{\sum_{i=1}^n M_{ii}^2}{n} + \left(1 - \frac{\sum_{i=1}^n Q_{ii}^2}{k}\right) \alpha_n\right] E^*[V_i^{*2}] E^*[\epsilon_i^{*2}] \\ \to^p \quad & (1 - \alpha) \sigma_{\epsilon\epsilon} \Sigma_{vv}\end{aligned}$$

where the first equality follows from the fact that $E^*[V_i^{*2} \epsilon_i^{*2}] = E^*[V_i^{*2}] E^*[\epsilon_i^{*2}]$ and $E^*[V_i^* \epsilon_i^*] = E^*[V_i^*] E^*[\epsilon_i^*] = 0$; the convergence in probability follows from

$$\frac{\sum_{i=1}^n M_{ii}^2}{n} + \left(1 - \frac{\sum_{i=1}^n Q_{ii}^2}{k}\right) \alpha_n \to 1 - \alpha$$

and $E^*[V_i^{*2}] \to^p \Sigma_{vv}, E^*[\epsilon_i^{*2}] \to^p \sigma_{\epsilon\epsilon}$.

Therefore, by Lemma A2 of Hansen, Hausman, and Newey (2008)

$$\frac{V^{*'} M_Z \epsilon^*}{\sqrt{n}} = \sum_{i=1}^n W_i^* - \alpha_n \sum_{i \neq j} \frac{V_i^* Q_{ij} \epsilon_j^*}{\sqrt{n}} \to^{d^*} N\left(0, (1 - \alpha) \sigma_{\epsilon\epsilon} \Sigma_{vv}\right)$$

in probability.

As has been show in Theorem 1 of Cattaneo, Jansson, and Newey (2012),

$$\sqrt{n} \left(\hat{\beta} - \beta\right) \to^d N\left(0, \Omega\right)$$

under Assumption 1 and 2, where $\Omega = \sigma_{\epsilon\epsilon} \Gamma^{-1}, \Gamma = (1 - \alpha) \Sigma_{vv}$.

Note that $\sqrt{n} \left(\hat{\beta}^* - \hat{\beta}\right) = \left(\widehat{\Gamma}^*\right)^{-1} \frac{X^{*'} M_Z \epsilon^*}{\sqrt{n}} = \left(\widehat{\Gamma}^*\right)^{-1} \frac{V^{*'} M_Z \epsilon^*}{\sqrt{n}}$. Using the results of Lemma A.2, we obtain by Contiunous Mapping Theorem

$$\sqrt{n}\left(\hat{\beta}^{*}-\hat{\beta}\right)\rightarrow^{d^{*}} N\left(0,\Omega\right)$$

in probability. The first result of the theorem then follows by applying Polya's Theorem, given that the normal distribution is everywhere continuous.

For the second result, note that

$$\begin{aligned}
\hat{s}^{*2} &= \frac{\hat{\epsilon}^{*\prime}\hat{\epsilon}^{*}}{n-k-1} \\
&= \frac{(y^{*}-X^{*}\hat{\beta}^{*})^{\prime}M_{Z}(y^{*}-X^{*}\hat{\beta}^{*})}{n-k-1} \\
&= \frac{\epsilon^{*\prime}M_{Z}\epsilon^{*}}{n-k-1}+\frac{(\hat{\beta}-\hat{\beta}^{*})^{\prime}V^{*\prime}M_{Z}\epsilon^{*}}{n-k-1}+\frac{\epsilon^{*\prime}M_{Z}V^{*}(\hat{\beta}-\hat{\beta}^{*})}{n-k-1}+\frac{(\hat{\beta}-\hat{\beta}^{*})^{\prime}V^{*\prime}M_{Z}V^{*}(\hat{\beta}-\hat{\beta}^{*})}{n-k-1} \\
&= \frac{\epsilon^{*\prime}M_{Z}\epsilon^{*}}{n-k-1}+O_{p^{*}}\left(\frac{1}{\sqrt{n}}\right)+O_{p^{*}}\left(\frac{1}{\sqrt{n}}\right)+O_{p^{*}}\left(\frac{1}{n}\right) \\
&\rightarrow^{p^{*}} \sigma_{\epsilon\epsilon}
\end{aligned}$$

where the third equality follows from the fact that $\frac{\epsilon^{*\prime}M_{Z}V^{*}}{n-k-1}\rightarrow^{p^{*}} 0$, $\frac{V^{*\prime}M_{Z}V^{*}}{n-k-1}\rightarrow^{p^{*}}\Sigma_{vv}$ which can be shown by using arguments similar to the proof for A.2, and from the fact that $\hat{\beta}^{*}-\hat{\beta}=O_{p^{*}}\left(\frac{1}{\sqrt{n}}\right)$; the convergence in probability follows from $\frac{\epsilon^{*\prime}M_{Z}\epsilon^{*}}{n-k-1}\rightarrow^{p^{*}}\sigma_{\epsilon\epsilon}$.

Moreover, we know from Theorem 1 of Cattaneo, Jansson, and Newey (2012) that

$$\widehat{\Omega}^{-1/2}\left(\hat{\beta}-\beta\right)\rightarrow^{d} N\left(0,1\right)$$

under Assumption 1 and 2.

Using the result of Lemma A.2 and the result that $\hat{s}^{2*}\rightarrow^{p^{*}}\sigma_{\epsilon\epsilon}$, we obtain by Continuous Mapping Theorem

$$\widehat{\Omega}^{*-1/2}\left(\hat{\beta}^{*}-\hat{\beta}\right)=\left(\hat{s}^{*2}\widehat{\Gamma}^{*-1}\right)^{-1/2}\left(\hat{\beta}^{*}-\hat{\beta}\right)\rightarrow^{d^{*}} N\left(0,1\right)$$

in probability. The second result then follows by applying Polya's Theorem. ■