A Simple Robust Procedure in Instrumental Variables Regression

Xiyu Jiao[*]
Department of Economics and Mansfield College
University of Oxford

November 23, 2018

Abstract

Due to the frequent concern that outliers may invalidate the empirical findings, in practical applications of instrumental variables regression the common practice is to first run ordinary two stage least squares and remove observations with residuals beyond a chosen cut-off value that classifies outliers. 2SLS is subsequently re-calculated with non-outlying observations, and this procedure is iterated until robust results are obtained. In this paper we analyze this simple robust algorithm asymptotically, then provide consistent estimation and valid inferential procedures for practical implementation given the cut-off value. Moreover, this paper provides asymptotic theory for setting the cut-off, which is chosen to control the gauge (proportion of outliers wrongly discovered). Asymptotics are derived under the null hypothesis that there is no contamination in the cross-sectional i.i.d. data. The established weak convergence result, involving empirical processes and fixed points, provides a starting point for statistical tests that assess model misspecification. Thus this paper also establishes the uniform and weak law for a new class of weighted and marked empirical processes, allowing for estimation errors of parameters in the structural IV equation.

Keywords: outlier robustness; instrumental variables; empirical processes; fixed point; uniform and weak convergence.

[*]Mansfield College, Mansfield Road, Oxford OX1 3TF, UK. E-mail: xiyu.jiao@mansfield.ox.ac.uk

1 Introduction

In applied economics research, all valid empirical findings should share one essential feature: that the results of analyses are not affected by a small portion of atypical observations to a nonnegligible degree; see Young (2017). Thus, empirical researchers are frequently concerned that their crucial findings are driven by a tiny set of outliers. For example, when estimating risk aversion coefficients, Toda and Walsh (2015) found that trimming a small sub-sample of observations drastically changes their estimates, making otherwise insignificant statistical inference significant.

In the empirical economics and finance literature, a recurrent critique is to ask if the paper examines the robustness of its conclusion, especially with respect to outlying observations, when classical or IVs regressions are applied. For instance, for De Long and Summers' (1991) paper investigating the effect of equipment investment on economic growth, Auerbach, Hassett, and Oliner (1994) questioned whether the positive capital effect is due to outliers. De Long and Summers (1994) agreed that they faced sample-selection issues and their findings are particularly sensitive to one outlier. Similarly in estimating the institutional impact on economic growth, Acemoglu, Johnson, and Robinson (2001) attracted extensive comments[1] on whether the validity of instruments by mortality rates is ruined by some outlying observations. Guthrie, Sokolowsky, and Wan's (2012) critique of Chhaochharia and Grinstein (2009) was that the removal of a few observations invalidates their empirical results, while a similar comment for Reinhart and Rogoff (2010) can be found in Herndon, Ash, and Pollin (2014).

In parallel to the empirical work, some recent theoretical papers in econometrics are also concerned about outlier contamination, particularly its effect on the performance of statistical inference. Taking Berenguer-Rico, Johansen and Nielsen (2018), Berenguer-Rico and Nielsen (2018), and Berenguer-Rico and Wilms (2018) as examples, they showed through simulations that outlying observations will substantially influence the outcome of specification checks which test if residuals are i.i.d. normal.

In response to the long-standing issue that potential outliers may drive empirical findings, the statistics literature[2] then designed some robust methods in regression which are less sensitive to outlying observations. In the context of instrumental variables regression, Young (2018) carried out a comprehensive bootstrap study of 1400 IV regressions in 32 papers published in journals of the American Economic Association, and then found that the two stage least squares estimator is more sensitive to outliers than ordinary least squares in classical regression. Thus, two types of robust methods (quantile and rank regression) were proposed and subsequently extended to the IV setting (see Honoré and Hu, 1994).

The quantile idea was first successfully introduced to regression by Koenker and Bassett (1978) to study the quantile effect (the whole distribution rather than just the mean effect) of treatment on the outcome of interest, and afterwards Chernozhukov and Hansen (2005) further generalized this method to simultaneous equations involving instruments. Meanwhile, within the local average treatment effect framework of Im-

[1] see for example Albouy (2012) and Acemoglu, Johnson, and Robinson (2012).

[2] see the thorough treatment in Rousseeuw and Leroy (1987) and Maronna, Martin, and Yohai (2006).

bens and Angrist (1994), Imbens and Newey (2009) provided another quantile approach through IVs based on Abadie, Angrist, and Imbens (2002)[3]. Compared to the quantile method, the rank regression by Adichie (1967) has been used much less often to perform IV in economics (see Honoré and Powell, 1994).

Although there are two existing IV robust methods in econometrics, the common practice in empirical IV research for outlier analysis is to first run ordinary two stage least squares and to find outliers by examining if residuals are higher than the selected cut-off. 2SLS is then re-computed with the non-outlying observations, and this procedure is iterated until the robust result is obtained. In practical applications of IVs regression, many economists in fact do apply this simple procedure to carry out robustness checks and compare the results from original ones relative to standard errors, see for instance De Long and Summers (1991, 1994), Auerbach, Hassett, and Oliner (1994), Acemoglu, Johnson, and Robinson (2001, 2012), Albouy (2012), Fabrizio, Rose, and Wolfram (2007), Toda and Walsh (2015), and Acemoglu et al. (2017).

Analyzing this common practice is significantly important for outlier analysis in empirical IV research, since asymptotic theory will give guidance on better implementing the algorithm to identify outliers, to obtain consistent and robust estimates, and to further make valid inference. However this heuristic approach lacks formal justification, and it is technically demanding to build up the probability tools (empirical processes) for establishing its statistical properties.

The main contribution of this paper is to analyse this simple robust IV algorithm asymptotically, then provide procedures for consistent estimation and valid inference given the cut-off value. Moreover, this paper contributes to providing asymptotic theory for setting the cut-off, like the tuning parameters in non-parametric econometrics, which should be chosen to control the gauge (the proportion of outliers falsely detected). Asymptotics in this paper are derived under the null hypothesis that there is no contamination in the cross-sectional i.i.d. data. The weak convergence result shown in this paper is sufficient to formalize IV outlier robustness checks frequently used by empirical researchers as a statistical test[4]. To carry out asymptotic analysis we generalize some recent results for marked and weighted empirical processes of residuals, which allow for variation in location, scale and quantile in the structural IV equation.

The heuristic procedure described above can be formalized as follows. Suppose we have independent and identically distributed data $\{(y_i, x_i, z_i)\}_{i=1}^n$ where the structural equation holds for outcome of interest and treatment variables as well as other characteristics such that $\{(y_i, x_i)\}_{i=1}^n$ follows

$$y_i = x_i'\beta + u_i, \quad i = 1, 2, \ldots, n,$$

while regressors and instruments $\{(x_i, z_i)\}_{i=1}^n$ satisfy the projection equation

$$x_i' = z_i'\Pi + r_i', \quad i = 1, 2, \ldots, n.$$

The error $(u_i/\sigma, \Sigma^{-1/2} r_i)$ scaled by its standard deviation is independent of filtration

$$\mathcal{F}_{i-1} = \sigma(z_1, \ldots, z_i, r_1, \ldots, r_{i-1}, u_1, \ldots, u_{i-1}),$$

[3]see discussion in Wüthrich (2018) for a connection between estimands of these two quantile models in IV.

[4]see Jiao and Pretis (2018) and Kaji (2018) whose tests however are mainly in the classical regression.

with the marginal and joint densities f_u, f_r, $\mathsf{f}_{u,r}$. Outliers are pairs of observations that do not conform with the structural equation so given an initial estimator $(\widetilde{\beta}, \widetilde{\sigma}^2)$ the robust algorithm decides which observations are non-outlying through

$$v_i = 1_{(|y_i - x_i'\widetilde{\beta}| \leq \widetilde{\sigma} c)},$$

where the choice of the cut-off c is related to the known reference density f_u. For those observations that are not outliers, re-run two stage least squares regression

$$\widehat{\beta} = (\widehat{\Pi}' \sum_{i=1}^{n} z_i z_i' v_i \widehat{\Pi})^{-1} (\widehat{\Pi}' \sum_{i=1}^{n} z_i y_i v_i), \qquad \widehat{\sigma}^2 = \varsigma^{-2} (\sum_{i=1}^{n} v_i)^{-1} \{\sum_{i=1}^{n} (y_i - x_i'\widehat{\beta})^2 v_i\},$$

where the location parameter in projection is estimated by

$$\widehat{\Pi} = (\sum_{i=1}^{n} z_i z_i' v_i)^{-1} (\sum_{i=1}^{n} z_i x_i' v_i).$$

The above one-step robust procedure can be iterated. Notice that ς^2 is the bias correction factor, that will be introduced in section §2, to ensure consistent estimation for variance σ^2. The iteration may also be initiated by a robust estimator. More simply two special algorithms heavily applied by empirical researchers are either starting with the full or split sample 2SLS. The idea of the latter was introduced in the empirical work of US food expenditure by Hendry (1999)[5].

Analysis of robust methods in the iterated one-step frame is by no means new. In the classical regression, one-step (Huber, L, and M) estimators have been considered by Bickel (1975), Ruppert and Carroll (1980), and Welsh and Ronchetti (2002). The idea of iterating one-step estimators can be found in Dollinger and Staudte (1991), who applied an influence function argument to demonstrate convergence of iteratively reweighted least squares with smooth weights. Notwithstanding we are interested in binary weights, the spirit of their argument still provides a guidance for our IV robust procedure. Other related work includes for example Cavaliere and Georgiev (2013), who explored the first order autoregression with infinite variance. For iterated one-step Huber-skip M and L-estimators, asymptotic theory has been recently established by Hendry, Johansen, and Santos (2008), Johansen and Nielsen (2009, 2013, 2016a, 2016b, 2018), and Jiao and Nielsen (2017). Maronna and Yohai (1997) and Hansen, Heaton, and Yaron (1996) also studied iterated (three stage least squares, GMM) estimators in the IVs regression but with the distinct concern on asymptotic efficiency when non-i.i.d. structural errors occur. The simple robust procedure studied in this paper actually mimics the iterated one-step Huber-skip M-estimator but in the IV setting. Thus, this paper extends asymptotics of iterated one-step Huber-skip M-estimators to the IVs regression where the exogeneity restriction on treatment variables breaks down, i.e. $\mathsf{E} x_i u_i \neq 0_{d_x}$.

The leading purpose of this work is to build an asymptotic theory for the simple robust procedure in the IVs regression which has been applied often in empirical research by many economists. In particular, we introduce the bias correction factor ς^2 for estimating σ^2 consistently and derive weak convergence theory for the algorithm, which

[5]also see Hendry and Doornik (2014).

together provide valid inference for structural parameters. Our analysis allows us to explore how variation in the cut-off c affects the robust IV algorithm, so this paper also gives guidance for selecting the threshold c indirectly from the gauge. Moreover, the weak convergence result in this paper is useful for testing whether the empirical findings are driven by outliers.

The argument involves a theory for a new class of weighted and marked empirical processes of residuals as well as the fixed point theory. The empirical process is defined from the generalized empirical distribution function

$$\widehat{\mathsf{F}}_{u,n}^{w,p}(a,b,c) = \frac{1}{n}\sum_{i=1}^{n} w_{in} u_i^p 1_{(u_i \leq \sigma c + n^{-1/2} ac + z_{in}' \Pi b + n^{-1/2} r_i' b)},$$

where the weights w_{in} are combinations of the normalized $\mathcal{F}_{i-1}$ measurable instruments z_{in} and u_i^p are the $\mathcal{F}_i$ adapted marks, while a, b represent the normalized estimation errors for σ, β. We derive asymptotic expansions that are uniform in a, b, c and allow for a near $n^{1/4}$ inefficiency in the estimation uncertainties a, b.

The empirical process literature dates back to Kolmogorov (1933) and Smirnov (1939) who proposed a type of goodness of fit tests that check whether the distribution under the null is well specified by comparing it to the empirical distribution function. To build asymptotics of Kolmogorov-Smirnov type test statistics, Doob (1949) first demonstrated weak convergence of empirical distribution function, and then Donsker (1952) established the empirical process central limit theorem to close a gap in Doob's proof. Because there exists the measurability issue in $D[0,1]$ space, the classical method applies the Skorokhod metric instead of uniform topology to avoid non-measurability[6]. Meanwhile, many researchers still equip $D[0,1]$ space with uniform metric but use the Hoffmann-Jørgensen[7] definition of weak convergence instead. With the new concept of weak convergence, they extend the classical theory to the empirical process indexed by VC class of sets or functions using the entropy and bracketing argument[8]. However, their work has not yet been extended to deal with the weights w_{in} and marks u_i^p appearing in our context.

Following the classical idea, Koul and Ossiander (1994) used the entropy argument to first build up the theory for the weighted empirical process in the autoregressive model[9]. Recently, empirical processes including both weights and marks have been analyzed by a series of papers, see Engler and Nielsen (2009), Johansen and Nielsen (2009, 2016a), Jiao and Nielsen (2017), Jiao (2018), Berenguer-Rico and Nielsen (2018), and Berenguer-Rico, Johansen, and Nielsen (2018). However, all these papers restrict attention to the classical regression model where regressors x_i are assumed to be orthogonal to errors u_i, i.e. $\mathsf{E} x_i u_i = 0_{d_x}$, so the filtration at $i-1$ can be simply constructed as $\mathcal{H}_{i-1} = \sigma(x_1, \ldots, x_i, u_1, \ldots, u_{i-1})$. But in the IV setting further dependence structure is involved between x_i and u_i or r_i and u_i making the problem complicated, and thus their argument

[6]see details in Billingsley (1968).

[7]a sequence of random elements X_n weakly converges to the limiting random element X if and only if $\mathsf{E}^* f(X_n) \to \mathsf{E} f(X)$ as $n \to \infty$ for any continuous and bounded function f where E^* denotes the outer expectation.

[8]see the summary in van der Vaart and Wellner (1996).

[9]also see Koul (2002) and Koul and Ling (2006).

does not go through and we have to reconstruct the martingale based on filtration $\mathcal{F}_{i-1}$ instead[10].

One of our contributions is to show uniform convergence and tightness for the empirical process constructed by $n^{1/2}\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$, which are essential for building up the weak convergence theory. Further we provide the linearization for the compensator of $n^{1/2}\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$. While $n^{1/2}$-convergence results are demonstrated for the process $n^{1/2}\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$, we finally establish n-convergence for a new type of empirical processes with marks r_i instead of u_i.

The paper proceeds as follows: §2 reviews a class of robust algorithms in the instrumental variables regression. Then, §3 presents the main asymptotic results for such robust procedures, while §4 provides uniform and weak law for a new class of the weighted and marked empirical processes with proofs in Appendix §A, B. Finally proofs of the main theorems in §3 follow in Appendix §C.

2 Model and a robust procedure

The instrumental variables (IV) regression with some notations is described first. It is widely known that two stage least squares estimator is sensitive to outliers. We review a robust procedure related to the iterated one-step Huber-skip M-estimator but in an IV setting. At last, define the concept of gauge to measure the rate of false detection of outlier detection algorithms. Then the cut-off value of such algorithms, like tuning parameters in nonparametrics, can be chosen indirectly from the gauge.

2.1 Model

Suppose in the cross-sectional settings we have independently and identically distributed data $\{(y_i,x_i,z_i)\}_{i=1}^n$ across individuals, where y_i is univariate and x_i, z_i are multivariate with dimension d_x, d_z and they are all demeaned so that $\mathsf{E}y_i=0$, $\mathsf{E}x_i=0_{d_x}$, $\mathsf{E}z_i=0_{d_z}$. Assume the data $\{(y_i,x_i)\}_{i=1}^n$ satisfies the structural equation

$$y_i = x_i'\beta + u_i, \quad i=1,2,\dots,n. \tag{2.1}$$

Structural errors $\{u_i\}_{i=1}^n$ are univariate i.i.d. random variables with scale σ so that u_i/σ has the known density $\mathsf{f}_u(y)$ and distribution function $\mathsf{F}_u(y)=\mathsf{P}(u_i/\sigma\le y)$ for $y\in\mathbb{R}$ with mean 0 and variance 1. Since in practice at least some elements of x_i are endogenous, i.e. $\mathsf{E}x_iu_i\neq 0_{d_x}$, instruments z_i are required for estimating the parameter

[10] In the classical setting where $\mathsf{E}x_iu_i=0_{d_x}$ holds such that the simple filtration $\mathcal{H}_{i-1}$ applies, Nielsen and Qian (2018) faced the related problem when investigating the empirical process of temporal differenced residuals

$$\widehat{\mathsf{H}}_n(a,b,c)=\frac{1}{n}\sum_{i=1}^n 1_{(\nabla u_i\le\sqrt{2}\sigma c+\sqrt{2}n^{-1/2}ac+\nabla x_{in}'b)},$$

where $\nabla u_i=u_i-u_{i+1}$, $\nabla x_{in}=x_{in}-x_{i+1,n}$, and x_{in} are normalized regressors due to its stochastic property, while a,b have the same meaning as in our setup. At time $i-1$ only x_{in} is $\mathcal{H}_{i-1}$ adapted while ∇u_i is independent of $\mathcal{H}_{i-1}$. Then dependence arises between ∇u_i and ∇x_{in} since u_i in ∇u_i is correlated with $x_{i+1,n}$ in ∇x_{in} at $\mathcal{H}_{i-1}$ although $\mathsf{E}x_iu_i=0_{d_x}$. But their issue is different from this paper since we relax independence between regressors x_i and errors u_i in the process.

$\beta \in \mathbb{R}^{d_x}$ consistently. The Hausman (1978) test was proposed to check if regressors x_i are uncorrelated with structural errors u_i based on the difference between ordinary least squares and two stage least squares estimators. Regressors and instruments $\{(x_i, z_i)\}_{i=1}^n$ satisfy the first stage regression

$$x_i' = z_i'\Pi + r_i', \quad i = 1, 2, \ldots, n. \tag{2.2}$$

Instruments z_i are assumed to be orthogonal to errors r_i in the first stage equation so $\mathsf{E} z_i r_i' = 0_{d_z \times d_x}$. Innovations $\{r_i\}_{i=1}^n$ are d_x-variate i.i.d. random vector with symmetric and positive definite dispersion matrix $\Sigma \in \mathbb{R}^{d_x \times d_x}$ so $\Sigma^{-1/2} r_i$ follows the density $\mathsf{f}_r(x)$ and distribution function $\mathsf{F}_r(x)$ for $x \in \mathbb{R}^{d_x}$ with mean 0_{d_x} and identity variance-covariance matrix I_{d_x}. The parameter Π lies in $\mathbb{R}^{d_z \times d_x}$ and we suppose $d_x \leq d_z$ meaning the number of regressors is less than or equal to instruments. Furthermore, to identify structural parameters β, instruments z_i need to be valid, i.e. $\mathsf{E} z_i u_i = 0_{d_z}$, and informative, i.e. rank condition for identification ($\operatorname{rank} \Pi = d_x$ and $\operatorname{rank} \mathsf{E} z_i z_i' = d_z$). In the over-identifying case, i.e. $d_x < d_z$, we can test validity of instruments z_i, see Sargan (1958) for two stage least squares or Hansen (1982) for GMM framework. In terms of examining if instruments z_i are informative, apply Cragg and Donald (1993) to test the rank condition for identification. Current literature focuses on the weak identification situation where in most practical applications the rank condition $\operatorname{rank} \Pi = d_x$ nearly fails. Staiger and Stock (1997) suggested the rule of thumb F-test critical values used to test if instruments z_i are weakly informative when only one endogenous variable is involved, see Stock and Yogo (2005) for multiple endogenous regressors. Under weak identification, usual statistical inference for β based on t or F tests does not work. Therefore some robust methods were then proposed to test statistical hypotheses on the structural parameters β when the informative condition of instruments z_i is in the margin of failure, see Anderson and Rubin (1949) or more powerful tests in Kleibergen (2002) and Moreira (2003).

To apply the martingale argument, we need to construct the filtration

$$\mathcal{F}_{i-1} = \sigma(z_1, \ldots, z_i, r_1, \ldots, r_{i-1}, u_1, \ldots, u_{i-1}) \tag{2.3}$$

such that u_i, r_i are $\mathcal{F}_i$ measurable and independent of $\mathcal{F}_{i-1}$ while z_i is adapted to $\mathcal{F}_{i-1}$. Notice in the IV setting u_i is correlated with r_i. Assume the scaled errors $(u_i/\sigma, \Sigma^{-1/2} r_i)$ have the joint density $\mathsf{f}_{u,r}(y, x)$ and distribution function $\mathsf{F}_{u,r}(y, x)$ for $y \in \mathbb{R}$, $x \in \mathbb{R}^{d_x}$. Note the joint distribution $\mathsf{f}_{u,r}, \mathsf{F}_{u,r}$ does also depend on the covariance $\Omega = \mathsf{Cov}(u_i/\sigma, \Sigma^{-1/2} r_i) = \mathsf{E}(\Sigma^{-1/2} r_i u_i/\sigma)$, however for simplicity $\Omega \in \mathbb{R}^{d_x}$ is suppressed in the notation of joint density. Further suppose the joint density can be decomposed into the conditional and marginal ones so $\mathsf{f}_{u,r}(y, x) = \mathsf{f}_{u|r}(y|x)\mathsf{f}_r(x) = \mathsf{f}_{r|u}(x|y)\mathsf{f}_u(y)$. Then denote the conditional expected value

$$\xi_y = \mathsf{E}(\Sigma^{-1/2} r_i | u_i/\sigma = y) = \int_{\mathbb{R}^{d_x}} x \mathsf{f}_{r|u}(x|y)(dx). \tag{2.4}$$

Notice $\xi_y \in \mathbb{R}^{d_x}$ is related to the covariance Ω between u_i/σ and $\Sigma^{-1/2} r_i$. In practice $\mathsf{f}_{u,r}$, $\mathsf{F}_{u,r}$ will often be assumed to be $(1 + d_x)$-variate normal, so the scaled error vector $(u_i/\sigma, \Sigma^{-1/2} r_i)$ follows

$$\begin{pmatrix} u_i/\sigma \\ \Sigma^{-1/2} r_i \end{pmatrix} \overset{\mathsf{D}}{=} \mathsf{N}\left\{ \begin{pmatrix} 0 \\ 0_{d_x} \end{pmatrix}, \begin{pmatrix} 1 & \Omega' \\ \Omega & I_{d_x} \end{pmatrix} \right\}. \tag{2.5}$$

As the marginal and conditional distribution of multivariate normal are still normally distributed, then $\Sigma^{-1/2} r_i | u_i/\sigma \sim \mathsf{f}_{r|u}$ follows normal $\mathsf{N}(\Omega u_i/\sigma, I_{d_x} - \Omega\Omega')$. Thus the conditional expectation in (2.4) has the form $\xi_y = \Omega y$. Denote

$$\zeta_y^+ = \xi_y + \xi_{-y}, \qquad \zeta_y^- = \xi_y - \xi_{-y}. \tag{2.6}$$

Under normality in (2.5), we have $\xi_{-y} = -\Omega y = -\xi_y$ so $\zeta_y^+ = 0_{d_x}$ and $\zeta_y^- = 2\xi_y = 2\Omega y$.

The robust procedure we analyzed detects outliers through checking if absolute standardised residuals in the structural equation (2.1) are beyond the chosen cut-off value c and then calculating the robust two stage least squares estimator from the non-outlying sample. This implicitly assumes symmetry of u_i/σ, while non-symmetry leads to specific forms of bias. We assume symmetry for f_u when analyzing the robust algorithm in §3, but not for the general empirical process results in §4. Let the absolute error $|u_i|/\sigma$ in (2.1) have a density $\mathsf{g}_u(y)$ and distribution function $\mathsf{G}_u(y) = \mathsf{P}(|u_i|/\sigma \le y)$ for $y > 0$. With a symmetry assumption, $\mathsf{G}_u(y) = 2\mathsf{F}_u(y) - 1$ and $\mathsf{g}_u(y) = 2\mathsf{f}_u(y)$. Define $\psi = \mathsf{G}_u(c)$ so the probability of exceeding the cut-off c is $\gamma = 1 - \psi$. Suppose the k-th moment of the density f_u exists, then introduce

$$\tau_k = \int_{-\infty}^{\infty} y^k \mathsf{f}_u(y) dy, \qquad \tau_k^c = \int_{-c}^{c} y^k \mathsf{f}_u(y) dy. \tag{2.7}$$

Thus $\tau_0^c = \psi$, $\tau_2 = 1$ while $\tau_k = \tau_k^c = 0$ for odd k when assuming symmetry for f_u. Define the conditional variance of u_i/σ given $(|u_i|/\sigma \le c)$ as

$$\varsigma_c^2 = \frac{\tau_2^c}{\psi} = \frac{\int_{-c}^{c} y^2 \mathsf{f}_u(y) dy}{\mathsf{P}(|u_i| \le \sigma c)}. \tag{2.8}$$

This will be used as a bias correction factor for the variance estimate computed from the selected non-outlying sample. With normality assumption (2.5), we have $u_i/\sigma \sim \mathsf{f}_u$ follows standard normal $\mathsf{N}(0,1)$ then $\tau_2^c = \psi - 2c\mathsf{f}_u(c)$, $\tau_4 = 3$, $\tau_4^c = 3\psi - 2c(c^2+3)\mathsf{f}_u(c)$.

For matrices M, choose the spectral norm $|M| = \max\{\text{eigen}(M'M)\}^{1/2}$ so that for vectors x then $|x|$ is the Euclidean norm. The spectral norm is compatible with respect to the Euclidean norm so $|Mx| \le |M||x|$. Notice (dM) represents the exterior product of all elements in the matrix M to describe the multivariate integral, see the chapter 2 of Muirhead (1982). For instance consider a vector $x \in \mathbb{R}^{d_x}$, then (dx) is the exterior product of some measures on $\mathbb{R}^{d_x}$.

2.2 Robust statistical algorithms to detect outliers

We first define an iterated version of robust two stage least squares estimators. Two specific examples with different initial estimates are then described.

An initial estimator for the parameter β is required to obtain all the residuals in the structural equation, so outliers are detected if the absolute value of standardised residuals in (2.1) are beyond the selected cut-off value. Running two stage least squares based on the non-outlying observations gives the robust estimator for β. This procedure can then be iterated and a formal version of the algorithm is defined in the following.

Algorithm 2.1. ***(Iterated 2SLS)***. *Choose a cut-off* $c > 0$.
1. Select initial estimators $\widehat{\beta}_c^{(0)}$, $(\widehat{\sigma}_c^{(0)})^2$ *and let* $m = 0$.
2. Define indicator variables for selecting non-outlying observations

$$v_{i,c}^{(m)} = 1_{(|y_i - x_i'\widehat{\beta}_c^{(m)}| \leq \widehat{\sigma}_c^{(m)} c)}. \tag{2.9}$$

3. Calculate least squares estimators for Π

$$\widehat{\Pi}_c^{(m+1)} = (\sum_{i=1}^n z_i z_i' v_{i,c}^{(m)})^{-1} (\sum_{i=1}^n z_i x_i' v_{i,c}^{(m)}). \tag{2.10}$$

4. Compute two stage least squares estimators for β *and* σ^2

$$\widehat{\beta}_c^{(m+1)} = (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i z_i' v_{i,c}^{(m)} \widehat{\Pi}_c^{(m+1)})^{-1} (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i y_i v_{i,c}^{(m)}), \tag{2.11}$$

$$(\widehat{\sigma}_c^{(m+1)})^2 = \varsigma_c^{-2} (\sum_{i=1}^n v_{i,c}^{(m)})^{-1} \{\sum_{i=1}^n (y_i - x_i'\widehat{\beta}_c^{(m+1)})^2 v_{i,c}^{(m)}\}. \tag{2.12}$$

5. Let $m = m + 1$ *and repeat 2, 3, and 4.*

Algorithm 2.1 is a similar procedure as iterated one-step Huber-skip M-estimators[11] but in the instrumental variables context instead of the classical regression setting. Algorithm 2.1 chooses the same sub-sample of non-outlying observations by the indicators $v_{i,c}^{(m)}$ for estimating Π in the first stage regression (2.2) and β, σ^2 in the structural equation (2.1). Notice the fact, that the sub-sample selected for (2.2) is the same as for (2.1), is essential for demonstrating consistency of estimators for the structural parameters β, σ^2, see the discussion in Remark C.2 (Appendix C.2). Furthermore, to estimate σ^2 consistently the bias correction factor ς_c^2 is required to adjust in (2.12), otherwise without ς_c^2 corrected in the variance estimator it will converge in probability to $\varsigma_c^2\sigma^2$ instead, see Remark C.3 (Appendix C.2).

In §3 we analyze this algorithm asymptotically and show how to choose the cut-off c indirectly from the concept of gauge. The algorithm could start with a robust estimator, while the so called Robustified 2SLS is initiated using the full sample 2SLS. The following algorithm formally defines Robustified 2SLS.

Algorithm 2.2. ***(Robustified 2SLS)***. *Choose a cut-off* $c > 0$.
1.1. Calculate least squares estimator for Π *based upon the whole sample*

$$\widehat{\Pi}_c^{(0)} = (\sum_{i=1}^n z_i z_i')^{-1} (\sum_{i=1}^n z_i x_i'). \tag{2.13}$$

1.2. Compute two stage least squares estimator for β *and* σ^2 *for the whole sample*

$$\widehat{\beta}_c^{(0)} = (\widehat{\Pi}_c^{(0)\prime} \sum_{i=1}^n z_i z_i' \widehat{\Pi}_c^{(0)})^{-1} (\widehat{\Pi}_c^{(0)\prime} \sum_{i=1}^n z_i y_i), \qquad (\widehat{\sigma}_c^{(0)})^2 = \frac{1}{n} \sum_{i=1}^n (y_i - x_i'\widehat{\beta}_c^{(0)})^2. \tag{2.14}$$

[11] one-step updated estimator mimics the Huber (1964) skip estimator, which has criterion function $\rho(y) = \min(y^2, c^2)/2$ as opposed to the Huber estimator with criterion function $\rho(y) = y^2/2$ for $|y| \leq c$ and $\rho(y) = c|y| - c^2/2$ otherwise, see Hampel et al. (1986, p. 104) and Jurečková, Sen, and Picek (2013, p. 175).

1.3. Choose $\widehat{\beta}_c^{(0)}$, $(\widehat{\sigma}_c^{(0)})^2$ *through* (2.14), *and let* $m = 0$.
2. Follow the step 2, 3, 4, and 5 in Algorithm 2.1.

Algorithm 2.2 is not very robust with respect to high leverage points when we have i.i.d. cross-sectional data. Thus, to deal with the high leverage points, another typical example of Algorithm 2.1 starting with the split sample 2SLS was first proposed but in the classical regression by Hendry (1999) in his empirical work[12]. This saturated idea is to divide the full sample into two sub-samples and use 2SLS calculated from each sub-sample to detect outliers in the other sub-sample. The following algorithm formally defines Saturated 2SLS.

Algorithm 2.3. *(**Saturated 2SLS**). Choose a cut-off* $c > 0$.
1.1. Split full sample into two sets $\mathcal{I}_j$, $j = 1, 2$ *of* n_j *observations where* $\sum_{j=1}^{2} n_j = n$.
1.2. Calculate least squares estimators for Π *based upon each sub-sample* $\mathcal{I}_j$ *for* $j = 1, 2$

$$\widehat{\Pi}_j = (\sum_{i \in \mathcal{I}_j} z_i z_i')^{-1} (\sum_{i \in \mathcal{I}_j} z_i x_i'). \tag{2.15}$$

1.3. Compute two stage least squares estimators for β *and* σ^2 *for sub-sample* $\mathcal{I}_j$, $j = 1, 2$

$$\widehat{\beta}_j = (\widehat{\Pi}_j' \sum_{i \in \mathcal{I}_j} z_i z_i' \widehat{\Pi}_j)^{-1} (\widehat{\Pi}_j' \sum_{i \in \mathcal{I}_j} z_i y_i), \qquad \widehat{\sigma}_j^2 = \frac{1}{n_j} \sum_{i \in \mathcal{I}_j} (y_i - x_i' \widehat{\beta}_j)^2. \tag{2.16}$$

1.4. Define the initial indicator variables for selecting non-outlying observations

$$v_{i,c}^{(0)} = 1_{(i \in \mathcal{I}_1)} 1_{(|y_i - x_i' \widehat{\beta}_2| \leq \widehat{\sigma}_2 c)} + 1_{(i \in \mathcal{I}_2)} 1_{(|y_i - x_i' \widehat{\beta}_1| \leq \widehat{\sigma}_1 c)}. \tag{2.17}$$

1.5. Compute $\widehat{\Pi}_c^{(1)}$, $\widehat{\beta}_c^{(1)}$, $(\widehat{\sigma}_c^{(1)})^2$ *using* (2.10), (2.11), (2.12) *with* $m = 0$, *and let* $m = 1$.
2. Follow the step 2, 3, 4, and 5 in Algorithm 2.1.

Algorithm 2.3 is possibly more robust than the Robustified Two Stage Least Squares when we have prior knowledge that outliers are located in a particular subset of the whole sample. Since the location of contaminated observations is unknown in most practical situations, the choice of the initial sets $\mathcal{I}_1$ and $\mathcal{I}_2$ should be iterated.

2.3 Gauge to measure the rate of false detection

Outlier detection algorithms have a positive probability to find outliers even when, in fact, the data generation process has no outliers. We evaluate the performance of such algorithms by the concept of gauge[13], which is the expected retention rate of falsely discovered outliers. This is a measure of type I error and it gives us an indirect

[12]see also Hendry, Johansen, and Santos (2008) and Hendry and Doornik (2014).

[13]Hoover and Perez (1999) originally introduced the idea of gauge in a simulation study of general-to-specific variable selection algorithms. The concept of gauge was formally proposed by Hendry and Santos (2010) as the expected retention rate of irrelevant regressors in the context of model selection algorithms. Doornik (2009) presented a comprehensive simulation study on the gauge for Algorithm 2.3 but only in the classical regression where x_i is assumed to be uncorrelated with u_i; also see Hendry and Doornik (2014).

way of choosing the cut-off value c. It is defined as follows. In the step m given the estimators $\widehat{\beta}_c^{(m)}$, $(\widehat{\sigma}_c^{(m)})^2$ and the cut-off c, the algorithms assign stochastic indicators $v_{i,c}^{(m)} = 1_{(|y_i - x_i'\widehat{\beta}_c^{(m)}| \leq \widehat{\sigma}_c^{(m)} c)}$ to all observations such as in (2.9) so that $v_{i,c}^{(m)} = 0$ when observation i is declared as an outlier, otherwise $v_{i,c}^{(m)} = 1$. When the model has no contamination, the sample and population gauge are

$$\widehat{\gamma}_c^{(m)} = \frac{1}{n}\sum_{i=1}^{n}(1 - v_{i,c}^{(m)}), \qquad \mathsf{E}\widehat{\gamma}_c^{(m)} = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(1 - v_{i,c}^{(m)}). \tag{2.18}$$

3 The main results

We start by listing assumptions, then asymptotics for Iterated 2SLS and weak convergence for Robustified 2SLS and Saturated 2SLS.

3.1 Assumptions

We list the sufficient assumptions for asymptotic theory of Algorithm 2.1, 2.2, 2.3. To simplify analysis, this section focuses on somewhat stronger conditions than they need to be. In section §4 on the one-sided empirical process, we will introduce some weaker assumptions. For example, we impose the symmetricity assumption on f_u in this section but not in section §4, though the idea of reasoning is the same and the results in §3 can be extended to the asymmetric case using the argument in Johansen and Nielsen (2009).

Since $\{(y_i, x_i, z_i)\}_{i=1}^n$ are i.i.d., instruments z_i can be normalized by the rate $n^{-1/2}$ so denote $z_{in} = n^{-1/2}z_i$, then we have $M_{zz,n} = \sum_{i=1}^n z_{in}z_{in}' = n^{-1}\sum_{i=1}^n z_i z_i' \overset{\mathsf{P}}{\to} \mathsf{E}z_i z_i' = M_{zz}$ by Law of Large Numbers assuming $\mathsf{E}|z_i|^2 < \infty$. Denote the infeasible ideal fitted value $\tilde{x}_i = \Pi' z_i$ while the feasible counterpart is attained by estimating Π consistently. The normalized $\tilde{x}_i$ is defined as $\tilde{x}_{in} = n^{-1/2}\tilde{x}_i = \Pi' z_{in}$ such that

$$M_{\tilde{x}\tilde{x},n} = \sum_{i=1}^{n}\tilde{x}_{in}\tilde{x}_{in}' = \Pi' M_{zz,n}\Pi \overset{\mathsf{P}}{\to} \Pi' M_{zz}\Pi = \mathsf{E}\tilde{x}_i\tilde{x}_i' = M_{\tilde{x}\tilde{x}}.$$

If $d_x \leq d_z$ and rank condition for identification hold so that $\operatorname{rank}\Pi = d_x$, $M_{zz} > 0$, then we have $M_{\tilde{x}\tilde{x}} > 0$ as well.

To carry out asymptotic analysis, the scaled error vector $(u_i/\sigma, \Sigma^{-1/2}r_i)$, instruments z_i, and the initial estimator $(\widetilde{\beta}, \widetilde{\sigma}^2)$ must satisfy the following conditions.

Assumption 3.1. *Let $\mathcal{F}_{i-1} = \sigma(z_1, \ldots, z_i, r_1, \ldots, r_{i-1}, u_1, \ldots, u_{i-1})$ be an increasing sequence of σ-fields so u_{i-1}, r_{i-1}, z_i are $\mathcal{F}_{i-1}$ measurable while r_i, u_i are independent of $\mathcal{F}_{i-1}$. Suppose $(u_i/\sigma, \Sigma^{-1/2}r_i)$ have continuously differentiable joint, conditional, and marginal densities $\mathsf{f}_{u,r}(y, x) = \mathsf{f}_{u|r}(y|x)\mathsf{f}_r(x) = \mathsf{f}_{r|u}(x|y)\mathsf{f}_u(y)$ which are positive on $y \in \mathbb{R}$, $x \in \mathbb{R}^{d_x}$. Assume $d_x \leq d_z$ and $\operatorname{rank}\Pi = d_x$. For $0 \leq \kappa < \eta \leq 1/4$, choose an integer $s \geq 2$ such that $2^{s-1} > 1 + (1/4 - \eta)(1 + d_x)$. Let $e = 1 + 2^{s+1}$. Suppose*
(i) the marginal density $\mathsf{f}_u(y)$ is symmetric and satisfies for $y \in \mathbb{R}$
(a) tail monotonicity: $y^e\mathsf{f}_u(y), |y^{e+1}\dot{\mathsf{f}}_u(y)|$ are decreasing for large y;

(ii) *the marginal density* $\mathsf{f}_r(x)$ *satisfies for* $x \in \mathbb{R}^{d_x}$

(a) *moments:* $\int_{x\in\mathbb{R}^{d_x}} |x|^4 \mathsf{f}_r(x)(dx) < \infty$*;*

(iii) *the joint and conditional densities* $\mathsf{f}_{u,r}(y,x), \mathsf{f}_{u|r}(y|x)$ *satisfy for* $y \in \mathbb{R}, x \in \mathbb{R}^{d_x}$

(a) *boundedness:* $\sup_{y\in\mathbb{R},x\in\mathbb{R}^{d_x}} |(1+y)y^{e-2}\mathsf{f}_{u|r}(y|x) + y^{e-1}\dot{\mathsf{f}}_{u|r,y}(y|x)| < \infty$*;*

(iv) *the instruments* z_i *satisfy*

(a) $M_{zz,n} = \sum_{i=1}^n z_{in} z_{in}' \overset{\mathsf{P}}{\to} M_{zz} > 0$*;*

(b) $\max_{1\le i\le n} |n^{1/2-\kappa} z_{in}| = \mathrm{O}_{\mathsf{P}}(1)$*;*

(c) $n^{-1}\mathsf{E}\sum_{i=1}^n |n^{1/2} z_{in}|^e = \mathsf{E}|n^{1/2} z_{in}|^e = \mathrm{O}(1)$*;*

(v) *the initial estimators* $(\widetilde{\beta}, \widetilde{\sigma}^2)$ *satisfy*

(a) $n^{1/2}(\widetilde{\beta} - \beta) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$*;*

(b) $n^{1/2}(\widetilde{\sigma}^2 - \sigma^2) = \mathrm{O}_{\mathsf{P}}(n^{1/4-\eta})$.

The conditions $(ia, iia, iiia)$ are satisfied in a range of situations. In particular (ia) is satisfied by the normal and t distribution; see Johansen and Nielsen (2016a, Example 3.1), while $(iia, iiia)$ holds for the normal distribution. Condition (iv) is standard for stationary instruments, see Johansen and Nielsen (2016a, Example 3.2). Condition (v) allows the standardised estimation errors to diverge at a rate of $n^{1/4-\eta}$ rather than being bounded in probability. In particular, $\eta = 1/4$ can be chosen for estimators with standard convergence rates.

Notice κ is not present in the moment condition $2^{s-1} > 1 + (1/4 - \eta)(1 + d_x)$, thus there is a trade-off between the convergence rate η of initial estimators and the required number s of moments for the density $\mathsf{f}_{u,r}$. In standard situations the normalized estimator is bounded so $\eta = 1/4$, then the moment condition to s reduces to $s = 2$. Otherwise when the initial estimator diverges at the rate η, then the number of moments s grows linearly with the dimension of the regressors d_x[14].

3.2 Asymptotics of Algorithm 2.1

This subsection shows asymptotic theory for Algorithms 2.1. Before establishing asymptotic properties of the iterated estimators for β, σ^2, we first demonstrate that the updated estimator for Π is consistent uniformly in $c \in [c_+, \infty)$ for a finite number $c_+ > 0$, given tightness of previous estimators for β, σ^2.

Theorem 3.1. *Consider Algorithm 2.1. Suppose Assumption 3.1*$(ia, iia, iiia, iv)$ *holds, and that* $n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$, $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ *are* $\mathrm{O}_{\mathsf{P}}(1)$ *for any* $m \in [0, \infty)$*. Then as* $n \to \infty$

$$\sup_{c_+\le c<\infty} |\widehat{\Pi}_c^{(m+1)} - \Pi| = \mathrm{o}_{\mathsf{P}}(1).$$

The proof of the above theorem uses empirical process theory studied in §4. Then with uniform consistency of location estimator for Π in the first stage regression (2.2) we build a one-step stochastic expansion of the updated estimators for structural parameters β, σ^2 in (2.1) in terms of its original estimators, kernels, and small remainder terms.

[14] also see discussion in Berenguer, Johansen, and Nielsen (2018, Remark 3.1, 3.2).

Theorem 3.2. *Consider Algorithm 2.1. Suppose Assumption 3.1(ia, iia, iiia, iv) holds, and that $n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$, $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are $\mathsf{O_P}(1)$ for any $m \in [0, \infty)$. Then as $n \to \infty$ and uniformly in $c \in [c_+, \infty)$*

$$n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta) = \frac{2c\mathsf{f}_u(c)}{\psi} n^{1/2}(\widehat{\beta}_c^{(m)} - \beta) + (M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)} + \mathsf{o_P}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c} n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma) + \frac{\sigma}{2\tau_2^c} n^{-1/2} \sum_{i=1}^{n} (\frac{u_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|u_i| \leq \sigma c)}$$

$$+ \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\tau_2^c} n^{1/2}(\widehat{\beta}_c^{(m)} - \beta) + \mathsf{o_P}(1).$$

Remark 3.1. *The above theorem explores expansion of Algorithm 2.1 (Iterated 2SLS) in the IV setting where $\mathsf{E}x_i u_i \neq 0_{d_x}$, which is the generalized result of iterated 1-step Huber-skip M-estimators in the classical regression; see Theorem 1 in Jiao and Nielsen (2017). To be specific, the expansion of β estimator only depends on its own estimation error in the previous step; see Proof of Theorem 3.2 (Appendix C.2) why $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ does not occur, and moreover if $\tilde{x}_{in}$ is replaced by x_{in} then it is the same as that in the classical regression. While the expansion of σ^2 estimator does also depend on the estimation error $n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$, this term disappears when $\mathsf{E}x_i u_i = 0_{d_x}$ back to the classical regression where we further have $\mathsf{E}r_i u_i = 0_{d_x}$ such that $\xi_c = 0_{d_x}$, $\zeta_c^+ = \zeta_c^- = 0_{d_x}$. Thus the second expansion degenerates to that in Jiao and Nielsen (2017).*

Remark 3.2. *Under normality (2.5) for $(u_i/\sigma, \Sigma^{-1/2} r_i)$ it follows $\zeta_c^- = 2\Omega c$ then*

$$\frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\tau_2^c} = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}\Omega'\Sigma^{1/2},$$

so for the expansion of variance estimator the fraction above appearing in the coefficient of $n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$ equals that of $n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$. Moreover, observe two expansions and find their autoregressive coefficients $2c\mathsf{f}_u(c)/\psi$, $c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)/\tau_2^c$. By Johansen and Nielsen (2013, Theorem 3.5), also see Jiao and Nielsen (2017, Theorem 2), Assumption 3.1(ia) implies they are strictly bounded by one so

$$\sup_{c_+ \leq c < \infty} \max\{|\frac{2c\mathsf{f}_u(c)}{\psi}|, |\frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}|\} < 1. \tag{3.1}$$

This in fact shows the spectral radius of the autoregressive coefficient matrix is smaller than one in the iterative system in Theorem 3.2, which is significant to establish the tightness and fixed point result, see the unit cycle boundedness condition (C.10) in the one-step stochastic expansion system (C.6), (C.7), (C.8) in Proof of Theorem 3.3 (Appendix C.2).

Assumption 3.1(v) with $\eta = 1/4$ corresponds to a standard convergence rate for the initial estimator. Theorem 3.2 provides an iterative equation between the updated and original estimators, while its autoregressive coefficient has a spectral radius strictly bounded by the unit cycle, see (3.1) in Remark 3.2. Thus a geometric argument and mathematical induction are then used to show $\widehat{\beta}_c^{(m)}$, $\widehat{\sigma}_c^{(m)}$ are tight in iteration $m \in [0, \infty)$ and in the cut-off value $c \in [c_+, \infty)$.

Theorem 3.3. *Consider Algorithm 2.1. Suppose Assumption 3.1 holds with $\eta = 1/4$. Then as $n \to \infty$*

$$\sup_{0\leq m<\infty}\sup_{c_+\leq c<\infty}|n^{1/2}(\widehat{\beta}_c^{(m)}-\beta)|+|n^{1/2}(\widehat{\sigma}_c^{(m)}-\sigma)| = \mathrm{O}_{\mathsf{P}}(1).$$

The above tightness result is required for building the weak convergence theory of iterated estimators for β, σ^2. Further combined with Theorem 3.1 a corollary follows, to demonstrate uniform consistency of $\widehat{\Pi}_c^{(m)}$ in m and c.

Corollary 3.1. *Consider Algorithm 2.1. Suppose Assumption 3.1 holds with $\eta = 1/4$. Then as $n \to \infty$*

$$\sup_{0\leq m<\infty}\sup_{c_+\leq c<\infty}|\widehat{\Pi}_c^{(m)}-\Pi| = \mathrm{o}_{\mathsf{P}}(1).$$

Since tightness results for iterated estimators of β, σ^2 have been established by Theorem 3.3, we can apply the one-step expansion in Theorem 3.2 recursively. Then for any $m \in [0,\infty)$ the stochastic expansions of $m+1$ step estimators are explored in terms of the initial estimators, kernels, and small remainder terms.

Theorem 3.4. *Consider Algorithm 2.1. Suppose Assumption 3.1 holds with $\eta = 1/4$. Then as $n \to \infty$ and uniformly in $c \in [c_+,\infty)$ we have for any $m \in [0,\infty)$*

$$n^{1/2}(\widehat{\beta}_c^{(m+1)}-\beta) = \varrho_{\beta\beta,c}^{(m+1)}n^{1/2}(\widehat{\beta}_c^{(0)}-\beta)+\varrho_{\beta\tilde{x}u,c}^{(m+1)}M_{\tilde{x}\tilde{x},n}^{-1}\sum_{i=1}^n\tilde{x}_{in}u_i1_{(|u_i|\leq\sigma c)}+\mathrm{o}_{\mathsf{P}}(1),$$

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)}-\sigma) = \varrho_{\sigma\sigma,c}^{(m+1)}n^{1/2}(\widehat{\sigma}_c^{(0)}-\sigma)+\varrho_{\sigma uu,c}^{(m+1)}n^{-1/2}\sum_{i=1}^n(\frac{u_i^2}{\sigma^2}-\varsigma_c^2)1_{(|u_i|\leq\sigma c)}$$
$$+\varrho_{\sigma\beta,c}^{(m+1)\prime}n^{1/2}(\widehat{\beta}_c^{(0)}-\beta)+\varrho_{\sigma\tilde{x}u,c}^{(m+1)\prime}M_{\tilde{x}\tilde{x},n}^{-1}\sum_{i=1}^n\tilde{x}_{in}u_i1_{(|u_i|\leq\sigma c)}+\mathrm{o}_{\mathsf{P}}(1),$$

where coefficients have expressions

$$\varrho_{\beta\beta,c}^{(m+1)} = \{\frac{2c\mathsf{f}_u(c)}{\psi}\}^{m+1},\qquad \varrho_{\beta\tilde{x}u,c}^{(m+1)} = \frac{\psi^{m+1}-\{2c\mathsf{f}_u(c)\}^{m+1}}{\psi^{m+1}\{\psi-2c\mathsf{f}_u(c)\}},$$
$$\varrho_{\sigma\sigma,c}^{(m+1)} = \{\frac{c(c^2-\varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}\}^{m+1},\qquad \varrho_{\sigma uu,c}^{(m+1)} = \sigma\frac{(\tau_2^c)^{m+1}-\{c(c^2-\varsigma_c^2)\mathsf{f}_u(c)\}^{m+1}}{2(\tau_2^c)^{m+1}\{\tau_2^c-c(c^2-\varsigma_c^2)\mathsf{f}_u(c)\}},$$
$$\varrho_{\sigma\beta,c}^{(m+1)} = \sum_{l=0}^m(\frac{2}{\psi})^{m-l}(\frac{c^2-\varsigma_c^2}{\tau_2^c})^{l+1}\frac{\{c\mathsf{f}_u(c)\}^{m+1}\Sigma^{1/2}\zeta_c^-}{2c},$$
$$\varrho_{\sigma\tilde{x}u,c}^{(m+1)} = [\frac{(\tau_2^c)^{m+1}-\{c(c^2-\varsigma_c^2)\mathsf{f}_u(c)\}^{m+1}}{(\tau_2^c)^m\{\tau_2^c-c(c^2-\varsigma_c^2)\mathsf{f}_u(c)\}}-\sum_{l=0}^m(\frac{2}{\psi})^{m-l}(\frac{c^2-\varsigma_c^2}{\tau_2^c})^l\{c\mathsf{f}_u(c)\}^m]$$
$$\times\frac{(c^2-\varsigma_c^2)\mathsf{f}_u(c)\Sigma^{1/2}\zeta_c^-}{2\tau_2^c\{\psi-2c\mathsf{f}_u(c)\}}.$$

For $m \in [0, \infty)$ the above expansion of $m+1$ step estimators generalizes the one-step expansion in Theorem 3.2. If $m = 0$ in Theorem 3.4, then we have

$$\varrho^{(1)}_{\beta\beta,c} = \frac{2c\mathsf{f}_u(c)}{\psi}, \qquad \varrho^{(1)}_{\beta\tilde{x}u,c} = \frac{1}{\psi}, \tag{3.2}$$
$$\varrho^{(1)}_{\sigma\sigma,c} = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}, \quad \varrho^{(1)}_{\sigma uu,c} = \frac{\sigma}{2\tau_2^c}, \quad \varrho^{(1)}_{\sigma\beta,c} = \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\Sigma^{1/2}\zeta_c^-}{2\tau_2^c}, \quad \varrho^{(1)}_{\sigma\tilde{x}u,c} = 0_{d_x},$$

so the expansion reduces to the one-step case shown in Theorem 3.2.

Initially the tight estimator is assumed to be available. This is iterated through the one-step equation presented in Theorem 3.2 with the spectral radius of its autoregressive coefficient bounded by one, see (3.1) in Remark 3.2, so when the iteration step becomes sufficiently large the iterated estimator converges in probability to the fixed point. To explore convergence in iteration, let $m \to \infty$ in Theorem 3.4 to get

$$\varrho^{(\infty)}_{\beta\beta,c} = 0, \qquad \varrho^{(\infty)}_{\beta\tilde{x}u,c} = \frac{1}{\psi - 2c\mathsf{f}_u(c)}, \tag{3.3}$$
$$\varrho^{(\infty)}_{\sigma\sigma,c} = 0, \qquad \varrho^{(\infty)}_{\sigma uu,c} = \frac{\sigma}{2\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\}},$$
$$\varrho^{(\infty)}_{\sigma\beta,c} = 0_{d_x}, \qquad \varrho^{(\infty)}_{\sigma\tilde{x}u,c} = \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\Sigma^{1/2}\zeta_c^-}{2\{\psi - 2c\mathsf{f}_u(c)\}\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\}},$$

then the fixed point $n^{1/2}(\widehat{\beta}_c^* - \beta) = n^{1/2}(\widehat{\beta}_c^{(\infty)} - \beta)$, $n^{1/2}(\widehat{\sigma}_c^* - \sigma) = n^{1/2}(\widehat{\sigma}_c^{(\infty)} - \sigma)$ follows.

Theorem 3.5. *Consider Algorithm 2.1. Suppose Assumption 3.1 holds with $\eta = 1/4$. Then for all $\epsilon, \delta > 0$ a pair $n_0 > 0, m_0 > 0$ exists, so for $n > n_0$ and $m > m_0$*

$$\mathsf{P}\{\sup_{c_+ \leq c < \infty} |n^{1/2}(\widehat{\beta}_c^{(m)} - \widehat{\beta}_c^*)| + |n^{1/2}(\widehat{\sigma}_c^{(m)} - \widehat{\sigma}_c^*)| > \delta\} < \epsilon,$$

where

$$n^{1/2}(\widehat{\beta}_c^* - \beta) = \frac{1}{\psi - 2c\mathsf{f}_u(c)} M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^n \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)},$$
$$n^{1/2}(\widehat{\sigma}_c^* - \sigma) = \frac{\sigma}{2\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\}} n^{-1/2} \sum_{i=1}^n (\frac{u_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|u_i| \leq \sigma c)}$$
$$+ \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\{\psi - 2c\mathsf{f}_u(c)\}\{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\}} M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^n \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)}.$$

Based on Theorem 3.3, if the initial estimator is bounded in a large compact set with large probability, then any iterated estimator takes values in the same compact set. The proof of Theorem 3.5 is to further argue that the deviation between the m-fold iterated estimator and the fixed point is the sum of two terms vanishing exponentially and in probability respectively when m and n tend to infinity.

Algorithm 2.1 mimics the Huber (1964) skip estimator in the IV context. By investigating Theorem 3.5 we reason that through infinite iterations the fixed point of the algorithm approximates the Huber-skip IV estimator in the sense that they have the same asymptotic expansion in terms of kernels and thus follow the same limiting distribution (see Theorem 3.8).

Similar to Remark 3.1, Theorems 3.4 and 3.5 extend the $(m+1)$-fold expansion and the fixed point result of the iterated 1-step Huber-skip M-estimator in classical regression to the IV setting where $\mathsf{E}x_i u_i \neq 0_{d_x}$ is allowed[15]. In practice we assume $(u_i/\sigma, \Sigma^{-1/2} r_i)$ follows normality (2.5) then $\xi_c = \Omega c$, $\zeta_c^+ = 0_{d_x}$, $\zeta_c^- = 2\Omega c$, so coefficients appearing in expansions in Theorems 3.4 and 3.5 can be further reduced as in Remark 3.2.

3.3 Weak convergence of Algorithm 2.2 and 2.3

Algorithms 2.2 and 2.3 are special versions of Algorithm 2.1 with different starting points. Their initial estimators are either the full sample or split sample two stage least squares which do not depend on the cut-off, and so satisfy the tightness property. Therefore, theorems and corollaries in the previous subsection apply for these two algorithms as well. Moreover, since Algorithms 2.2 and 2.3 start with two stage least squares estimators whose statistical properties are well known, the asymptotic distribution can then be established for these two robust procedures.

When researchers carry out empirical work using instrumental variables regression, their interest mainly lies in making inference on β whereas they are only concerned about the consistency of estimators for σ^2, Π. Theorem 3.3 and Corollary 3.1 have already shown that iterated estimators of σ^2, Π are consistent, so in order to perform inference on structural parameters in (2.1) we next build distributional theory for the iterated estimator of β in Algorithms 2.2 and 2.3.

Choosing the distinct cut-off value c in the interval $[c_+, \infty)$, we then get a process $\mathbb{G}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ for any $m \in [0, \infty)$. A weak convergence theory for $\mathbb{G}_n^{(m+1)}$ follows from a finite dimensional convergence derived by the expansion in Theorem 3.4 and tightness in Theorem 3.3 (see Theorem 13.1 in Billingsley, 1968). The iterated estimator of β as a process is asymptotically approximated by the Gaussian process[16].

We first analyze Algorithm 2.2 by providing asymptotics for the full sample two stage least squares estimator $\widetilde{\beta}$ defined through its initial estimate $\widehat{\beta}_c^{(0)}$ in (2.14).

Lemma 3.1. *Consider the full sample two stage least squares*

$$\widetilde{\beta} = (\widetilde{\Pi}' \sum_{i=1}^n z_i z_i' \widetilde{\Pi})^{-1} (\widetilde{\Pi}' \sum_{i=1}^n z_i y_i), \quad \text{where} \quad \widetilde{\Pi} = (\sum_{i=1}^n z_i z_i')^{-1} (\sum_{i=1}^n z_i x_i').$$

Suppose Assumption 3.1(ia, iia, iva, ivc) holds. Then as $n \to \infty$

$$n^{1/2}(\widetilde{\beta} - \beta) = M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^n \tilde{x}_{in} u_i + \mathsf{o}_{\mathsf{P}}(1).$$

Furthermore we have

$$n^{1/2}(\widetilde{\beta} - \beta) \xrightarrow{\mathsf{D}} \mathsf{N}(0_{d_x}, \sigma^2 M_{\tilde{x}\tilde{x}}^{-1}).$$

[15]replace $\tilde{x}_{in}$ by x_{in} and set $\mathsf{E}x_i u_i = 0_{d_x}$ so $\mathsf{E}r_i u_i = 0_{d_x}$, we then have $\xi_c = 0_{d_x}$, $\zeta_c^+ = \zeta_c^- = 0_{d_x}$ thus all results degenerate to those shown in Jiao and Nielsen (2017).

[16]see the definition of Gaussian processes in Adler and Taylor (2009, p. 27).

The next step is to establish the weak convergence theory for the process $\mathbb{G}_n^{(m+1)}$ of the iterated β estimator where $m \in [0, \infty)$.

Theorem 3.6. *Consider Algorithm 2.2. Suppose Assumption 3.1(ia, iia, iiia, iv) holds. Denote the process $\mathbb{G}_n^{(m+1)}(c) = n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ for $c \in [c_+, \infty)$ and $m \in [0, \infty)$. Then as $n \to \infty$ we have*

$$\mathbb{G}_n^{(m+1)}(c) = \varrho_{\beta\beta,c}^{(m+1)} M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i + \varrho_{\beta\tilde{x}u,c}^{(m+1)} M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i 1_{(|u_i| \le \sigma c)} + \mathsf{o}_\mathsf{P}(1),$$

where $\varrho_{\beta\beta,c}^{(m+1)}$, $\varrho_{\beta\tilde{x}u,c}^{(m+1)}$ are defined in Theorem 3.4. Furthermore $\mathbb{G}_n^{(m+1)}$ weakly converges to a zero mean Gaussian process $\mathbb{G}^{(m+1)}$ with variance given as

$$\mathsf{Var}\{\mathbb{G}^{(m+1)}(c)\} = \{(\varrho_{\beta\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta\beta,c}^{(m+1)} \varrho_{\beta\tilde{x}u,c}^{(m+1)} + \tau_2^c (\varrho_{\beta\tilde{x}u,c}^{(m+1)})^2\} \sigma^2 M_{\tilde{x}\tilde{x}}^{-1}.$$

The one-step updated estimator from the full sample 2SLS is of particular interest in Algorithm 2.2. To explore its asymptotics let $m = 0$ in Theorem 3.6 such that $\varrho_{\beta\beta,c}^{(1)} = 2c\mathsf{f}_u(c)/\psi$ and $\varrho_{\beta\tilde{x}u,c}^{(1)} = \psi^{-1}$ as in (3.2), then the corollary below follows.

Corollary 3.2. *Consider Algorithm 2.2. Suppose Assumption 3.1(ia, iia, iiia, iv) holds. Denote the process $\mathbb{G}_n^{(1)}(c) = n^{1/2}(\widehat{\beta}_c^{(1)} - \beta)$ for $c \in [c_+, \infty)$. Then as $n \to \infty$ we have*

$$\mathbb{G}_n^{(1)}(c) = \frac{2c\mathsf{f}_u(c)}{\psi} M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i + (M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i 1_{(|u_i| \le \sigma c)} + \mathsf{o}_\mathsf{P}(1).$$

Furthermore $\mathbb{G}_n^{(1)}$ weakly converges to a zero mean Gaussian process $\mathbb{G}^{(1)}$ with variance given as

$$\mathsf{Var}\{\mathbb{G}^{(1)}(c)\} = \frac{4c^2\mathsf{f}_u^2(c) + 4\tau_2^c c\mathsf{f}_u(c) + \tau_2^c}{\psi^2} \sigma^2 M_{\tilde{x}\tilde{x}}^{-1}.$$

Then we analyze Algorithm 2.3 and find asymptotically it is equivalent to Algorithm 2.2 although they start with distinct initial estimates when implementing Algorithm 2.1.

Theorem 3.7. *Consider Algorithm 2.3 where $n_1 = n_2 = n/2$. Suppose Assumption 3.1(ia, iia, iiia, iv) holds for each sub-sample set $\mathcal{I}_1$, $\mathcal{I}_2$. Then as $n \to \infty$ and for any $m \in [0, \infty)$, the process of the estimator $n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta)$ for $c \in [c_+, \infty)$ has the same asymptotic expansion as for Algorithm 2.2 and thus weakly converges to the identical Gaussian process reported in Theorem 3.6.*

The proof involves checking whether the expansion of the first step updated estimator for Algorithm 2.3 is the same as Algorithm 2.2 using arguments in Theorems 3.1 and 3.2. Since the two algorithms are iterated from the split half or full sample 2SLS respectively, they have the same asymptotics if their $\widehat{\beta}_c^{(1)}$ perform identically when $n \to \infty$.

Finally, when $n \to \infty$ and $m \to \infty$, the process $\mathbb{G}_n^{(m)}$ uniformly converges to the limiting process $\mathbb{G}^{(\infty)} = \mathbb{G}^*$ in probability. Set $\varrho_{\beta\beta,c}^{(\infty)} = 0$, $\varrho_{\beta\tilde{x}u,c}^{(\infty)} = 1/\{\psi - 2c\mathsf{f}_u(c)\}$ as in (3.3), then Theorems 3.6 and 3.7 immediately show the weak convergence of the fixed point process $\mathbb{G}^*$ for Algorithms 2.1, 2.2, and 2.3.

Theorem 3.8. *Consider Algorithm 2.1 with tight initial estimators, such as Algorithm 2.2, 2.3. Suppose Assumption 3.1(ia, iia, iiia, iv) holds. When $m \to \infty$ let the fixed point process $\mathbb{G}_n^*(c) = n^{1/2}(\widehat{\beta}_c^* - \beta)$ for $c \in [c_+, \infty)$ as in Theorem 3.5. Then as $n \to \infty$ the process $\mathbb{G}_n^*$ has the asymptotic expansion shown in Theorem 3.5 and thus weakly converges to a zero mean Gaussian process $\mathbb{G}^*$ with variance given as*

$$\mathsf{Var}\{\mathbb{G}^*(c)\} = \frac{\tau_2^c}{\{\psi - 2c\mathsf{f}_u(c)\}^2}\sigma^2 M_{\tilde{x}\tilde{x}}^{-1}.$$

Consider Algorithms 2.2 and 2.3. First notice that weak convergence in Theorems 3.6 and 3.7 implies pointwise convergence, so for any $c \in [c_+, \infty)$, $m \in [0, \infty)$, and as $n \to \infty$ it holds

$$n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta) \overset{\mathsf{D}}{\to} \mathsf{N}[0_{d_x}, \{(\varrho_{\beta\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta\beta,c}^{(m+1)} \varrho_{\beta\tilde{x}u,c}^{(m+1)} + \tau_2^c(\varrho_{\beta\tilde{x}u,c}^{(m+1)})^2\}\sigma^2 M_{\tilde{x}\tilde{x}}^{-1}], \quad (3.4)$$

where $\varrho_{\beta\beta,c}^{(m+1)}$, $\varrho_{\beta\tilde{x}u,c}^{(m+1)}$ are defined in Theorem 3.4. Lemma 3.1 shows the asymptotic distribution of the ordinary two stage least squares estimator $\widetilde{\beta}$. Now we can compare the efficiency of the robust estimator $\widehat{\beta}_c^{(m+1)}$ with respect to the non-robust estimator $\widetilde{\beta}$ under the null of no outliers. Efficiency is then

$$\begin{aligned}\text{efficiency}(\widehat{\beta}_c^{(m+1)}, \widetilde{\beta}) &= \{\mathsf{asVar}(\widehat{\beta}_c^{(m+1)})\}^{-1}\{\mathsf{asVar}(\widetilde{\beta})\} \\ &= \{(\varrho_{\beta\beta,c}^{(m+1)})^2 + 2\tau_2^c \varrho_{\beta\beta,c}^{(m+1)} \varrho_{\beta\tilde{x}u,c}^{(m+1)} + \tau_2^c(\varrho_{\beta\tilde{x}u,c}^{(m+1)})^2\}^{-1} I_{d_x}. \quad (3.5)\end{aligned}$$

We are interested in efficiency comparison for two special cases: when $m = 0$ and $m \to \infty$, so substituting (3.2) and (3.3) into (3.5) we then have

$$\text{efficiency}(\widehat{\beta}_c^{(1)}, \widetilde{\beta}) = \frac{\psi^2}{4c^2\mathsf{f}_u^2(c) + 4\tau_2^c c\mathsf{f}_u(c) + \tau_2^c} I_{d_x}, \quad (3.6)$$

$$\text{efficiency}(\widehat{\beta}_c^*, \widetilde{\beta}) = \frac{\{\psi - 2c\mathsf{f}_u(c)\}^2}{\tau_2^c} I_{d_x}. \quad (3.7)$$

Using the standard normal density for f_u, we then plot (3.6), (3.7) in terms of critical values $c \in [c_+, \infty)$ in Figure 3.1.

Asymptotics in this paper are derived under the null hypothesis of no outliers where Algorithms 2.2 and 2.3 have a probability to falsely trim the non-outlying observations given the cut-off c. Thus robust estimates have a higher asymptotic variance relative to the ordinary 2SLS, and so the efficiency plot lies in the interval $(0, 1)$ as in Figure 3.1. When critical values c become larger, the chance of wrongly detecting outliers tends to be smaller. Subsequently, robust estimates will have the smaller variance, and their efficiency become larger and closer to one, and vice versa if c decreases. First observe the efficiency plot of the infinite step estimator $\widehat{\beta}_c^*$ in Figure 3.1, and find it is consistent with what we discussed. Then check the solid line of the first step estimator $\widehat{\beta}_c^{(1)}$, and find that its efficiency tends to one when $c \to 0$ as well as when $c \to \infty$. It is intuitive that for large c the efficiency increases as c increases. However we also find for small c the efficiency increases as c decreases. This is due to the fact that the updated estimator $\widehat{\beta}_c^{(1)}$ is asymptotically equivalent to 2SLS when $c \to 0$ as stated in Remark 3.3.

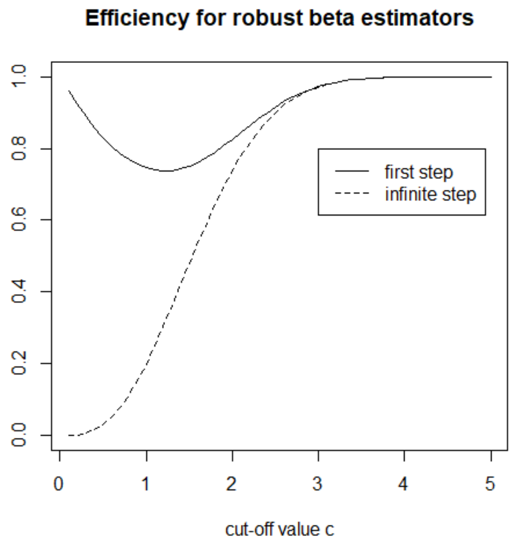

Figure 3.1: The efficiency of $\widehat{\beta}_c^{(1)}$ (solid line) and $\widehat{\beta}_c^{*}$ (dashed line) with respect to the two stage least squares $\widetilde{\beta}$ for f_u equal to the standard normal density and $c \in [0.1, 5]$.

Remark 3.3. *Let $m = 0$ and $n \to \infty$, Theorem 3.4 gives the equation*

$$n^{1/2}(\widehat{\beta}_c^{(1)} - \beta) = \frac{2c\mathsf{f}_u(c)}{\psi} n^{1/2}(\widetilde{\beta} - \beta) + (M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)} + \mathrm{o}_{\mathsf{P}}(1),$$

uniformly in $c \in [c_+, \infty)$. Note that $\widetilde{\beta}$ is 2SLS so $\widehat{\beta}_c^{(1)}$ is its updated estimator. Suppose f_u follows the standard normal then $\psi = 2\mathsf{F}_u(c) - 1$, $\tau_2^c = \psi - 2c\mathsf{f}_u(c)$, and $\dot{\mathsf{f}}_u(c) = -c\mathsf{f}_u(c)$. Let $c_+ > 0$ be sufficiently small so that we can investigate the situation where $c \to 0$. Lemma C.6 (Appendix C.3) shows the weak limit of the kernel term such that

$$(M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)} \xrightarrow{\mathsf{D}} \mathsf{N}(0_{d_x}, \frac{\tau_2^c}{\psi^2} \sigma^2 M_{\tilde{x}\tilde{x}}^{-1}).$$

By L'Hôpital's rule, we have $\tau_2^c/\psi^2 \to 0$ as $c \to 0$ so the kernel term vanishes. Again when $c \to 0$, L'Hôpital's rule gives $2c\mathsf{f}_u(c)/\psi \to 1$, thus $n^{1/2}(\widehat{\beta}_c^{(1)} - \beta)$ is asymptotically equivalent to $n^{1/2}(\widetilde{\beta} - \beta)$ indicating an efficiency of unity.

In fact, check the efficiency formula given by (3.6). When $c \to 0$, use $\tau_2^c/\psi^2 \to 0$ and $2c\mathsf{f}_u(c)/\psi \to 1$ to get $\psi^2/\{4c^2\mathsf{f}_u^2(c) + 4\tau_2^c c\mathsf{f}_u(c) + \tau_2^c\} \to 1$, then find $\mathrm{efficiency}(\widehat{\beta}_c^{(1)}, \widetilde{\beta}) \to 1$.

Figure 3.1 implies Algorithms 2.2 and 2.3 do not lose too much efficiency compared to the ordinary 2SLS if the cut-off value c is appropriately chosen under the null of no outliers, whereas these two procedures will provide the more robust result under the alternative when there is data contamination.

Johansen and Nielsen (2009, 2013) built up the asymptotic distributions for the one-step Huber-skip M-estimator and its infinite iteration either starting from full sample or split sample least squares in the classical setting where $\mathsf{E}x_i u_i = 0_{d_x}$[17]. We then find the distributional results for Algorithms 2.2 and 2.3 are suprisingly the same as the iterated one-step Huber-skip M-estimators if $\tilde{x}_{in}$ is replaced by x_{in} even though dependence is allowed between regressors and errors in the IV regression. This is due to the fact that the stochastic expansion for the updated β estimator only depends on its own previous step, not on the σ estimator (see Theorem 3.2 and Remark 3.1). Furthermore this result is because ζ_c^+ defined in (2.6, §2.1) does not appear in the β expansion since the data is demeaned such that $\mathsf{E}z_i = 0_{d_z}$[18].

When making inference on the structural parameters β using Algorithms 2.2 and 2.3, we need to estimate σ^2 appeared in the asymptotic variance shown in (3.4), given the cut-off value c and the density f_u. By introducing the bias correction factor ς_c^2 in (2.8, §2.1), then σ^2 can be estimated consistently under the null when there is no data contamination (see Theorem 3.3 and Remark C.3).

Nuisance parameters ξ_c, ζ_c^+, ζ_c^- measure the dependence between x_i and u_i (between r_i and u_i) in the IV setup. They are significantly important to evaluate the degree of endogeneity in the structural equation, but difficult to estimate in practice. However, performing inference on β does not require estimating these nuisance parameters, since they all vanish asymptotically as discussed. Therefore, in practical applications Algorithms 2.2 and 2.3 can be easily implemented not only to identify outliers and to obtain a robust and consistent estimator, but also to carry out valid inference.

4 Weighted and marked empirical process

Consider the weighted and marked empirical distribution function

$$\widehat{\mathsf{F}}_{u,n}^{w,p}(a,b,c) = \frac{1}{n}\sum_{i=1}^{n} w_{in} u_i^p 1_{(u_i \leq \sigma c + n^{-1/2}ac + z_{in}'\Pi b + n^{-1/2}r_i' b)}, \tag{4.1}$$

with $\mathcal{F}_{i-1}$ adapted weights w_{in} and $\mathcal{F}_i$ measurable marks u_i^p. The filtration $\mathcal{F}_{i-1}$ is generated by $(z_1,\ldots,z_i,r_1,\ldots,r_{i-1},u_1,\ldots,u_{i-1})$ so $z_{in} \in \mathcal{F}_{i-1}$ while $u_i, r_i \in \mathcal{F}_i$ are independent of $\mathcal{F}_{i-1}$. The observable data $\{(y_i,x_i,z_i)\}_{i=1}^n$ has i.i.d. structure in this paper, so $a \in \mathbb{R}$, $b \in \mathbb{R}^{d_x}$ represent normalized estimation errors $\widetilde{a} = n^{1/2}(\widetilde{\sigma} - \sigma)$, $\widetilde{b} = n^{1/2}(\widetilde{\beta} - \beta)$, and $c \in \mathbb{R}$ is the quantile, while σ, Π, Σ are true parameters for the variance of the structural error, for the location coefficient in the first stage regression, and for the first stage error respectively. Let normalized instruments $z_{in} = n^{-1/2}z_i$ such that $M_{zz,n} = \sum_{i=1}^n z_{in}z_{in}' = n^{-1}\sum_{i=1}^n z_i z_i'$ converges in probability to $\mathsf{E}z_i z_i' = M_{zz}$ by Law of Large Numbers when $\mathsf{E}|z_i|^2 < \infty$. Our interest focuses on weights w_{in} given as either of 1, $n^{1/2}z_{in} = z_i$, $n z_{in} z_{in}' = z_i z_i'$ and p as either of 0, 1, 2. To form the empirical process, introduce the compensator

$$\overline{\mathsf{F}}_{u,n}^{w,p}(a,b,c) = \frac{1}{n}\sum_{i=1}^{n} w_{in} \mathsf{E}_{i-1} u_i^p 1_{(u_i \leq \sigma c + n^{-1/2}ac + z_{in}'\Pi b + n^{-1/2}r_i' b)}, \tag{4.2}$$

[17]also see Johansen and Nielsen (2016b) for the asymptotic distribution of the $(m+1)$-step estimator where $m \in [0,\infty)$.

[18]see the details in Proof of Theorem 3.2, Appendix C.2.

where $\mathsf{E}_{i-1}(\cdot) = \mathsf{E}(\cdot|\mathcal{F}_{i-1})$. Note that $\overline{\mathsf{F}}^{1,0}_{u,n}(0,0,c) = \mathsf{F}_u(c) = \mathsf{P}(u_i \leq \sigma c)$.

We embed these processes in the space $D[0,1]$ that are processes continuous from the right and with limits of left, where the space is endowed with the Skorokhod metric. We do this as follows. The indicator $1_{(u_i \leq \sigma c)}$ and the distribution function $\mathsf{F}_u(c)$ can be defined as 0 or 1 when c takes the values $-\infty$ and ∞ respectively. We can then define quantiles $c_\psi = \mathsf{F}_u^{-1}(\psi)$ for $0 \leq \psi \leq 1$. Correspondingly we can continously extend the definition of the weighted and marked empirical distribution function and its compensator by chosing $\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,-\infty) = \overline{\mathsf{F}}^{w,p}_{u,n}(a,b,-\infty) = 0$ while we then have $\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,\infty) = n^{-1}\sum_{i=1}^n w_{in} u_i^p$ and $\overline{\mathsf{F}}^{w,p}_{u,n}(a,b,\infty) = n^{-1}\sum_{i=1}^n w_{in}\mathsf{E}_{i-1}u_i^p$. We now define the empirical process for $0 \leq \psi \leq 1$

$$\mathbb{F}^{w,p}_{u,n}(a,b,c_\psi) = n^{1/2}\{\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c_\psi) - \overline{\mathsf{F}}^{w,p}_{u,n}(a,b,c_\psi)\}. \tag{4.3}$$

In the following first present assumptions, and results then follow for the one-sided process and finally extend to the absolute case.

4.1 Assumptions

In this section, the density f_u is not necessarily symmetric. The conditions listed below are weaker than Assumption 3.1.

Assumption 4.1. *Let $\mathcal{F}_{i-1} = \sigma(z_1,\ldots,z_i,r_1,\ldots,r_{i-1},u_1,\ldots,u_{i-1})$ be an increasing sequence of σ-fields so u_{i-1}, r_{i-1}, z_i, w_{in} are $\mathcal{F}_{i-1}$ measurable while r_i, u_i are independent of $\mathcal{F}_{i-1}$. Suppose $(u_i/\sigma, \Sigma^{-1/2}r_i)$ have continuously differentiable joint, conditional, and marginal densities $\mathsf{f}_{u,r}(y,x) = \mathsf{f}_{u|r}(y|x)\mathsf{f}_r(x) = \mathsf{f}_{r|u}(x|y)\mathsf{f}_u(y)$ which are positive on $y \in \mathbb{R}$, $x \in \mathbb{R}^{d_x}$. Let p,η,κ be given so $p \in \mathbb{N}_0$, $0 \leq \kappa < \eta \leq 1/4$. Choose $0 < \nu \leq 1$, $s \in \mathbb{N}_0$ such that*

$$2^{s-1} > 1 + (1/4 - \eta)(1 + d_x). \tag{4.4}$$

Suppose
(i) the marginal density $\mathsf{f}_u(y)$ satisfies for $y \in \mathbb{R}$
  (a) moments: $\int_{-\infty}^{\infty} |y|^{2^s p/\nu}\mathsf{f}_u(y)dy < \infty$;
  (b) boundedness: $\sup_{y\in\mathbb{R}} |y^{2^s p+1}\mathsf{f}_u(y) + y^{2^s p+2}\dot{\mathsf{f}}_u(y)| < \infty$;
  (c) smoothness: a $C_{\mathrm{H}} \in \mathbb{N}$ exists so that for all $\epsilon > 0$

$$\frac{\sup_{y\geq\epsilon}(1+y^{2^s p})\mathsf{f}_u(y)}{\inf_{0\leq y\leq\epsilon}(1+y^{2^s p})\mathsf{f}_u(y)} \leq C_{\mathrm{H}}, \qquad \frac{\sup_{y\leq-\epsilon}(1+y^{2^s p})\mathsf{f}_u(y)}{\inf_{-\epsilon\leq y\leq 0}(1+y^{2^s p})\mathsf{f}_u(y)} \leq C_{\mathrm{H}};$$

(ii) the marginal density $\mathsf{f}_r(x)$ satisfies for $x \in \mathbb{R}^{d_x}$
  (a) moments: $\int_{x\in\mathbb{R}^{d_x}} |x|^4\mathsf{f}_r(x)(dx) < \infty$;
(iii) the joint and conditional densities $\mathsf{f}_{u,r}(y,x), \mathsf{f}_{u|r}(y|x)$ satisfy for $y \in \mathbb{R}, x \in \mathbb{R}^{d_x}$
  (a) boundedness: $\sup_{y\in\mathbb{R},x\in\mathbb{R}^{d_x}} |(1+y)y^{2^s p-1}\mathsf{f}_{u|r}(y|x) + y^{2^s p}\dot{\mathsf{f}}_{u|r,y}(y|x)| < \infty$;
(iv) the instruments z_i satisfy
  (a) $\max_{1\leq i\leq n}|n^{1/2-\kappa}z_{in}| = \mathrm{O}_{\mathsf{P}}(1)$;
(v) the weights w_{in} satisfy
  (a) $n^{-1}\mathsf{E}\sum_{i=1}^n |w_{in}|^{2^s}(1+|n^{1/2}z_{in}|) = \mathrm{O}(1)$;
  (b) $n^{-1}\sum_{i=1}^n |w_{in}|(1+|n^{1/2}z_{in}|^2) = \mathrm{O}_{\mathsf{P}}(1)$.

Remark 4.1. *Assumption 3.1(ia, iia, iiia, ivb, ivc) implies Assumption 4.1 with $s \geq 2$ satisfying (4.4) when w_{in} is either of 1, $n^{1/2}z_{in} = z_i$, $nz_{in}z'_{in} = z_i z'_i$ and p is either of 0, 1, 2. Details are given in Lemma B.1 in the appendix.*

4.2 Asymptotic results for empirical process

We present three asymptotic results. The first theorem shows that the estimation error for the scale and regression parameters is negligible uniformly in the quantile.

Theorem 4.1. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1 holds with $\nu = 1$ and $s \geq 2$ satisfying (4.4). Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\leq\psi\leq 1} \sup_{|a|,|b|\leq n^{1/4-\eta}B} |\mathbb{F}^{w,p}_{u,n}(a, b, c_\psi) - \mathbb{F}^{w,p}_{u,n}(0, 0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(1).$$

The proof involves a chaining argument. For this, we apply an iterated martingale inequality, see Lemma A.3, to explore the tail behaviour of the maximum of a family of martingales. We can, however, split this argument into two parts. First, we keep b fixed and consider variation in a, c_ψ. Then we keep a fixed and consider variation in b, c_ψ. Combining these two arguments will finally finish the proof.

The second result provides a linearization of the compensator.

Theorem 4.2. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia, ib, iia, iiia, vb) holds with $\nu = 1$, $s = 0$. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\leq\psi\leq 1} \sup_{|a|,|b|\leq n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{F}}^{w,p}_{u,n}(a, b, c_\psi) - \overline{\mathsf{F}}^{w,p}_{u,n}(0, 0, c_\psi)\} - \mathcal{B}_{\mathsf{F}_u,n}(a, b, c_\psi)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}),$$

where the bias term, with $\xi_{c_\psi} = \mathsf{E}(\Sigma^{-1/2}r_i|u_i/\sigma = c_\psi)$, is defined as

$$\mathcal{B}_{\mathsf{F}_u,n}(a, b, c_\psi) = \sigma^{p-1}c^p_\psi \mathsf{f}_u(c_\psi)n^{-1/2}\sum_{i=1}^n w_{in}(n^{-1/2}ac_\psi + n^{-1/2}\xi'_{c_\psi}\Sigma^{1/2}b + z'_{in}\Pi b).$$

Since x_i and u_i are correlated so for u_i and r_i, this dependent structure requires a more intricate analysis for the compensator. Similar as the proof for uniform convergence in Theorem 4.1, we proceed by first considering variation in a, c_ψ while $b = 0$. We then approximate the compensator by combining with the argument for variation in b, c_ψ while $a = 0$.

Finally, we argue the empirical process $\mathbb{F}^{w,p}_{u,n}(0, 0, c_\psi)$ is tight when viewed as a sequence in n of processes on $D[0, 1]$. By Billingsley (1968, Theorem 13.2), two conditions need to be checked. First, it holds by construction that $\mathbb{F}^{w,p}_{u,n}(0, 0, c_0) = 0$ where $c_0 = \mathsf{F}_u^{-1}(0) = -\infty$. Second, the next theorem shows that the modulus of continuity is small. The proof uses a dyadic argument and then apply an iterated martingale exponential inequality, see Lemma A.4, to explore the tail probability of the maximum of a family of martingales.

Theorem 4.3. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia, va) holds with $0 < \nu < 1$, $s = 2$. Then for all $\epsilon > 0$*

$$\lim_{\phi\downarrow 0}\limsup_{n\to\infty} \mathsf{P}\{\sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |\mathbb{F}^{w,p}_{u,n}(0, 0, c_{\psi^\dagger}) - \mathbb{F}^{w,p}_{u,n}(0, 0, c_\psi)| > \epsilon\} \to 0.$$

The empirical distribution functions $\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$ and $\widehat{\Pi}^{(m+1)}_c$ involve in the updated estimators $\widehat{\beta}^{(m+1)}_c$, $(\widehat{\sigma}^{(m+1)}_c)^2$, see (2.11), (2.12). While $n^{1/2}$-convergence results have been established for the empirical process $n^{1/2}\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$ by Theorem 4.1, 4.2, 4.3, consistency for the estimator of Π is then required to build up the asymptotic distribution theory for the iterated estimators of β, σ^2. One new product moment appearing in $\widehat{\Pi}^{(m+1)}_c$ but not included in $\widehat{\mathsf{F}}^{w,p}_{u,n}(a,b,c)$ is $n^{-1}\sum_{i=1}^n z_i r_i' v^{(m)}_{i,c}$, see (2.9), (2.10). Thus for proving consistency of the Π estimator we need the following n-convergence theorem for a new class of empirical process with the weight $n^{1/2}z_{in} = z_i$ and mark r_i' instead of mark u_i in the previous case.

Theorem 4.4. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1 holds with $\nu = 1$ and $s \geq 2$ satisfying (4.4). Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\leq\psi\leq 1}\ \sup_{|a|,|b|\leq n^{1/4-\eta}B} |n^{-1/2}\sum_{i=1}^n z_{in} r_i' \{1_{(u_i \leq \sigma c_\psi + n^{-1/2}ac_\psi + z_{in}'\Pi b + n^{-1/2}r_i'b)} - 1_{(u_i\leq \sigma c_\psi)}\}| = \mathrm{o}_{\mathsf{P}}(1).$$

The proof first considers uniform convergence for the empirical process without weights $n^{1/2}z_{in} = z_i$ and marks r_i. Since we establish n-convergence result instead of with the normalization $n^{1/2}$, the Hölder inequality is then used to prove Theorem 4.4 by adding the weights and marks into the process with only indicators.

4.3 A result for the two-sided empirical process

Estimators in Algorithm 2.1, 2.2, 2.3 involves indicators depending on the absolute value of residuals of structural errors. We therefore present some results for a class of two-sided weighted and marked empirical processes.

Define the weighted and marked absolute empirical distribution function

$$\widehat{\mathsf{G}}^{w,p}_{u,n}(a,b,c) = \frac{1}{n}\sum_{i=1}^n w_{in} u_i^p 1_{(|u_i - z_{in}'\Pi b - n^{-1/2}r_i'b| \leq \sigma c + n^{-1/2}ac)}. \tag{4.5}$$

We suppose a such that $\sigma + n^{-1/2}a > 0$, in which case it suffices to consider $c \geq 0$. This restriction on a is satisfied when choosing a as $\widetilde{a} = n^{1/2}(\widetilde{\sigma} - \sigma)$ so that $\sigma + n^{-1/2}\widetilde{a} = \widetilde{\sigma} > 0$. Introduce the compensator of $\widehat{\mathsf{G}}^{w,p}_{u,n}(a,b,c)$

$$\overline{\mathsf{G}}^{w,p}_{u,n}(a,b,c) = \frac{1}{n}\sum_{i=1}^n w_{in}\mathsf{E}_{i-1} u_i^p 1_{(|u_i - z_{in}'\Pi b - n^{-1/2}r_i'b| \leq \sigma c + n^{-1/2}ac)}. \tag{4.6}$$

Note that $\overline{\mathsf{G}}^{1,0}_{u,n}(0,0,c) = \mathsf{G}_u(c) = \mathsf{P}(|u_i| \leq \sigma c)$. Similar as defining the one-sided empirical process $\mathbb{F}^{w,p}_{u,n}(a,b,c_\psi)$ on the space $D[0,1]$ equipped with the Skorokhod metric, let $c_\psi = \mathsf{G}_u^{-1}(\psi)$ for $0 \leq \psi \leq 1$ so $c_\psi \geq 0$ and the absolute empirical process is

$$\mathbb{G}^{w,p}_{u,n}(a,b,c_\psi) = n^{1/2}\{\widehat{\mathsf{G}}^{w,p}_{u,n}(a,b,c_\psi) - \overline{\mathsf{G}}^{w,p}_{u,n}(a,b,c_\psi)\}. \tag{4.7}$$

We can now derive asymptotic theory for the absolute empirical process from Theorems 4.1, 4.2, 4.3, 4.4. These results are presented under more restrictive Assumption

3.1, where the distribution f_u of structural error is symmetric, see Remark 4.1. In this section, we only consider w_{in} chosen as 1, $n^{1/2}z_{in} = z_i$, $nz_{in}z_{in}' = z_i z_i'$ and p as 0, 1, 2.

The first theorem states the absolute processes with estimation errors converge to the one without estimation errors uniformly in the quantile.

Theorem 4.5. *Let $c_\psi = \mathsf{G}_u^{-1}(\psi)$. Suppose Assumption 3.1(ia, iia, iiia, ivb, ivc) holds. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\le\psi\le 1} \sup_{|a|,|b|\le n^{1/4-\eta}B} |\mathbb{G}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{G}_{u,n}^{w,p}(0, 0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(1).$$

Then we provide the first-order approximation to the absolute compensator.

Theorem 4.6. *Let $c_\psi = \mathsf{G}_u^{-1}(\psi)$. Suppose Assumption 3.1(ia, iia, iiia, ivc) holds. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\le\psi\le 1} \sup_{|a|,|b|\le n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{G}}_{u,n}^{w,p}(a, b, c_\psi) - \overline{\mathsf{G}}_{u,n}^{w,p}(0, 0, c_\psi)\} - \mathcal{B}_{\mathsf{G}_u,n}(a, b, c_\psi)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}),$$

where the bias term, with $\xi_{c_\psi} = \mathsf{E}(\Sigma^{-1/2} r_i | u_i/\sigma = c_\psi)$, is defined as

$$\mathcal{B}_{\mathsf{G}_u,n}(a, b, c_\psi) = \sigma^{p-1} c_\psi^p \mathsf{f}_u(c_\psi) n^{-1/2} \sum_{i=1}^n w_{in}[\{1 + (-1)^p\} n^{-1/2} a c_\psi + n^{-1/2}\{\xi_{c_\psi} - (-1)^p \xi_{-c_\psi}\}'\Sigma^{1/2} b + \{1 - (-1)^p\} z_{in}'\Pi b].$$

The next theorem extends the tightness result to the case of empirical processes with absolute indicators.

Theorem 4.7. *Let $c_\psi = \mathsf{G}_u^{-1}(\psi)$. Suppose Assumption 3.1(ia, ivc) holds. Then for all $\epsilon > 0$*

$$\lim_{\phi\downarrow 0} \limsup_{n\to\infty} \mathsf{P}\{\sup_{0\le\psi\le\psi^\dagger\le 1: \psi^\dagger - \psi \le \phi} |\mathbb{G}_{u,n}^{w,p}(0, 0, c_{\psi^\dagger}) - \mathbb{G}_{u,n}^{w,p}(0, 0, c_\psi)| > \epsilon\} \to 0.$$

While the above theorems deal with $n^{1/2}$-convergence results, the final result considers n-convergence for a new type of absolute processes used to prove consistency of the estimator for Π in the first stage regression (2.2).

Theorem 4.8. *Let $c_\psi = \mathsf{G}_u^{-1}(\psi)$. Suppose Assumption 3.1(ia, iia, iiia, ivb, ivc) holds. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\le\psi\le 1} \sup_{|a|,|b|\le n^{1/4-\eta}B} |n^{-1/2} \sum_{i=1}^n z_{in} r_i' \{1_{(|u_i - z_{in}'\Pi b - n^{-1/2} r_i' b| \le \sigma c_\psi + n^{-1/2} a c_\psi)} - 1_{(|u_i| \le \sigma c_\psi)}\}| = \mathrm{o}_{\mathsf{P}}(1).$$

The proofs of empirical processes are given in Appendix B, but first we discuss a metric on $\mathbb{R}$ and provide some useful inequalities applied repeatedly in the chaining argument in Appendix A. Finally, the proofs of the main results follow in Appendix C.

5 Conclusion and discussion

Since most empirical analyses are affected by atypical observations, when many applied economists apply IVs regressions they first run ordinary 2SLS and use the resulting residuals to find non-outlying observations. They then re-run 2SLS on this subset, subsequently iterating this procedure until they obtain robust results.

In this paper we analyze this simple robust algorithm asymptotically, and provide the consistent estimation and valid inferential procedures given the cut-off value. We concentrate on cross-sectional i.i.d. data in this paper, though our analysis can be extended to time series, whether stationary, deterministic trend, or unit root. The results are all derived under the null hypothesis that there are no outliers in the model. Since in this case there is still a positive probability to find outliers using such algorithms, we then propose the concept of gauge - the expected retention rate of falsely discovered outliers - to evaluate their performance (see 2.18, §2.3).

Future work should focus on building asymptotic theory for the gauge, to give empirical researchers an indirect way of choosing the cut-off. Furthermore, some specification tests could be devised to check whether outliers truly exist with respect to the reference model (2.1, 2.2), as in Jiao and Pretis (2018). Similarly, a statistical test could also be proposed to check whether there is significant difference between robust and ordinary 2SLS, as in Kaji (2018).

It is well known that the first order asymptotic approximation is fragile in some small finite sample situations. Therefore, it would be of interest to carry out simulation studies to evaluate the finite sample performance of the results in this paper. Likewise it would be of interest to extend these results to situations where outliers are actually present in the data generating process. Possible scenarios include single outliers, clusters of outliers, level shifts, symmetric or non-symmetric outliers. In such situations, we would analyze the potency, which is the retention rate for relevant outliers.

A A metric on $\mathbb{R}$ and some inequalities

The asymptotic theory uses a chaining argument. This involves a partitioning of the quantile axis using a metric, which is presented first. Then follows some preliminary inequalities including an iterated exponential martingale inequality.

For $x, y \in \mathbb{R}$ define the function

$$J_{i,p}(x,y) = (\frac{u_i}{\sigma})^p\{1_{(u_i/\sigma\leq y)} - 1_{(u_i/\sigma\leq x)}\}. \tag{A.1}$$

Our interest focus on $J_{i,p}(x,y)$ of order 2^s with $s \in \mathbb{N}$. For $y, z \in \mathbb{R}$ note $z^{2^s p}$ is non-negative since $2^s p$ is even for $p \in \mathbb{N}_0$ and $s \in \mathbb{N}$, so introduce a positive and increasing function

$$\mathsf{H}_s(y) = \int_{-\infty}^{y} (1 + z^{2^s p})\mathsf{f}_u(z)dz. \tag{A.2}$$

The derivative of this function is $\dot{\mathsf{H}}_s(y) = (1 + y^{2^s p})\mathsf{f}_u(y)$. Then, denote the constant

$$H_s = \mathsf{H}_s(\infty) = \int_{-\infty}^{\infty} (1 + z^{2^s p})\mathsf{f}_u(z)dz, \tag{A.3}$$

which is assumed to be finite. Selection of the specific $s \in \mathbb{N}$ will be more clear in proofs of the empirical process results. The intuition of $\mathsf{H}_s(y)$ is obtained through setting $p = 0$ so that $\mathsf{H}_s(y) = 2\mathsf{F}_u(y)$, $\dot{\mathsf{H}}_s(y) = 2\mathsf{f}_u(y)$ and $H_s = 2$. Therefore, $\mathsf{H}_s(y)$ is the generalization of the distribution $\mathsf{F}_u(y) \sim u_i/\sigma$. Define the metric d on $x, y \in \mathbb{R}$ as

$$\mathsf{d}(x, y) = |\mathsf{H}_s(x) - \mathsf{H}_s(y)|. \tag{A.4}$$

Notice the metric d is totally bounded on $\mathbb{R}$ since $\mathsf{d}(-\infty, \infty) = H_s < \infty$, and it is a special case of the pseudo metric proposed by Koul and Ossiander (1994). For $x \leq y$ and $1 \leq q \leq s$,

$$0 \leq \mathsf{E}_{i-1} J_{i,p}(x, y)^{2^q} = \mathsf{E} J_{i,p}(x, y)^{2^q} < \mathsf{H}_s(y) - \mathsf{H}_s(x) = \mathsf{d}(x, y), \tag{A.5}$$

as $|z|^q < 1 + |z|^s$ for $0 \leq q \leq s$.

In the context of chaining, partition the range of $\mathsf{H}_s(c)$ into K intervals of equal size H_s/K. In other words, use the metric d to partition $\mathbb{R}$ into K intervals by endpoints

$$-\infty = c_0 < c_1 < \cdots < c_{K-1} < c_K = \infty, \tag{A.6}$$

with $c_{-k} = c_0$ for $k \in \mathbb{N}$ so that $\mathsf{d}(c_{k-1}, c_k) = H_s/K$ for $1 \leq k \leq K$.

The next inequality provides a tightness type result for the metric d, which will be used to prove the tightness property of empirical process $\mathbb{F}^{w,p}_{u,n}(0, 0, c_\psi)$.

Lemma A.1. *(Johansen and Nielsen, 2016a, Lemma B.3). Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia) holds with $0 < \nu < 1$, $s \in \mathbb{N}_0$. Then there exist C_ν, $0 < \phi_0 < 1$ so that for $0 \leq \phi \leq \phi_0$ it follows $\sup_{0 \leq \psi \leq 1-\phi} \mathsf{d}(c_\psi, c_{\psi+\phi}) \leq C_\nu \phi^{1-\nu}$.*

The next lemma is the bias correction for differences in expectations involving two dependent random variables, which will be used to linearize the compensator.

Lemma A.2. *Let (Y, X) be a random vector with dimension $1 + d_x$ with the joint density $\mathsf{m}(y, x)$ for $y \in \mathbb{R}, x \in \mathbb{R}^{d_x}$. Suppose the joint density decomposes into the conditional and marginal densities as $\mathsf{m}(y, x) = \mathsf{m}_{Y|X}(y|x)\mathsf{m}_X(x) = \mathsf{m}_{X|Y}(x|y)\mathsf{m}_Y(y)$ and its derivative for y exists so $\dot{\mathsf{m}}_y(y, x) = \dot{\mathsf{m}}_{Y|X,y}(y|x)\mathsf{m}_X(x)$. Let $b_1 \in \mathbb{R}$, $b_2 \in \mathbb{R}^{d_x}$, and $p \in \mathbb{N}_0$. Assume $\mathsf{E}|X|^2 < \infty$, then*

$$\begin{aligned}
&\sup_{c \in \mathbb{R}} |\mathsf{E} Y^p \{1_{(Y \leq c + b_1 + X' b_2)} - 1_{(Y \leq c)}\} - c^p \mathsf{m}_Y(c)\{b_1 + \mathsf{E}(X' b_2 | Y = c)\}| \\
\leq& (\frac{1}{2} b_1^2 + |b_1||b_2| \mathsf{E}|X| + \frac{1}{2}|b_2|^2 \mathsf{E}|X|^2) \sup_{y \in \mathbb{R}, x \in \mathbb{R}^{d_x}} |p y^{p-1} \mathsf{m}_{Y|X}(y|x) + y^p \dot{\mathsf{m}}_{Y|X,y}(y|x)|.
\end{aligned}$$

Proof of Lemma A.2. Let $\mathcal{E} = \mathsf{E} Y^p \{1_{(Y \leq c + b_1 + X' b_2)} - 1_{(Y \leq c)}\}$, and write the expectation as an integral

$$\mathcal{E} = \iint_{y \in \mathbb{R}, x \in \mathbb{R}^{d_x}} y^p 1_{(c \leq y \leq c + b_1 + x' b_2)} \mathsf{m}(y, x) dy(dx) = \int_{x \in \mathbb{R}^{d_x}} \int_c^{c + b_1 + x' b_2} y^p \mathsf{m}(y, x) dy(dx).$$

Apply Taylor expansion at point c to the inner integral to obtain

$$\int_c^{c + b_1 + x' b_2} y^p \mathsf{m}(y, x) dy = (b_1 + x' b_2) c^p \mathsf{m}(c, x) + \frac{1}{2}(b_1 + x' b_2)^2 \{p \tilde{c}^{p-1} \mathsf{m}(\tilde{c}, x) + \tilde{c}^p \dot{\mathsf{m}}_y(\tilde{c}, x)\},$$

where $|\tilde{c} - c| \leq |b_1 + x'b_2|$. Then $\mathcal{E}$ is approximated by the first order term $\mathcal{E}_1$ with the remainder term $\mathcal{E}_2$ left so $\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2$ where

$$\mathcal{E}_1 = \int_{\mathbb{R}^{d_x}} (b_1 + x'b_2) c^p \mathsf{m}(c, x)(dx),$$
$$\mathcal{E}_2 = \frac{1}{2} \int_{\mathbb{R}^{d_x}} (b_1 + x'b_2)^2 \{p\tilde{c}^{p-1} \mathsf{m}(\tilde{c}, x) + \tilde{c}^p \dot{\mathsf{m}}_y(\tilde{c}, x)\}(dx).$$

Analyze the first order term $\mathcal{E}_1$ to get

$$\mathcal{E}_1 = c^p \mathsf{m}_Y(c) b_1 + c^p \mathsf{m}_Y(c) \int_{\mathbb{R}^{d_x}} x'b_2 \mathsf{m}_{X|Y}(x|c)(dx) = c^p \mathsf{m}_Y(c)\{b_1 + \mathsf{E}(X'b_2|Y = c)\}.$$

For the remainder term $\mathcal{E}_2 = \mathcal{E} - \mathcal{E}_1$, apply the triangle inequality so

$$|\mathcal{E}_2| \leq \frac{1}{2} \int_{\mathbb{R}^{d_x}} \{b_1^2 + 2|b_1 x'b_2| + (x'b_2)^2\} |p\tilde{c}^{p-1} \mathsf{m}_{Y|X}(\tilde{c}|x) + \tilde{c}^p \dot{\mathsf{m}}_{Y|X,y}(\tilde{c}|x)| \mathsf{m}_X(x)(dx).$$

As $|b_1 x'b_2| \leq |b_1||b_2||x|$ and $(x'b_2)^2 \leq |x|^2|b_2|^2$, so

$$|\mathcal{E}_2| \leq (\frac{1}{2} b_1^2 + |b_1||b_2|\mathsf{E}|X| + \frac{1}{2}|b_2|^2 \mathsf{E}|X|^2) \sup_{y \in \mathbb{R}, x \in \mathbb{R}^{d_x}} |py^{p-1} \mathsf{m}_{Y|X}(y|x) + y^p \dot{\mathsf{m}}_{Y|X,y}(y|x)|. \blacksquare$$

The chaining argument involves the tail behaviour of the maximum of a family of martingales which can be controlled using the iterated exponential martingale inequality from Johansen and Nielsen (2016a). It builds on an exponential martingale inequality derived by Bercu and Touati (2008, Theorem 2.1) which relaxes the need of boundedness for martingale difference series but required in Freedman (1975) inequality. The following are two special cases of iterated exponential martingale inequalities, where the number of elements in the martingale family is increasing and where it is fixed. The first inequality will be used to demonstrate the uniform convergence result for empirical processes, while the second is for handling the tightness property.

Lemma A.3. *(Johansen and Nielsen, 2016a, Theorem 5.2) For l so $1 \leq l \leq L$, let $z_{l,i}$ be $\mathcal{F}_i$ adapted satisfying $\mathsf{E} z_{l,i}^{2^s} < \infty$ for some $s \in \mathbb{N}$. Let $D_q = \max_{1 \leq l \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{l,i}^{2^q}$ for $1 \leq q \leq s$. Suppose, for some $\varsigma \geq 0$, $\lambda > 0$, that $L = \mathrm{O}(n^\lambda)$ and $\mathsf{E} D_q = \mathrm{O}(n^\varsigma)$ for $1 \leq q \leq s$. If $\upsilon > 0$ is chosen such that*
(i) $\varsigma < 2\upsilon$;
(ii) $\varsigma + \lambda < \upsilon 2^s$;
then, we have for all $\kappa > 0$

$$\lim_{n \to \infty} \mathsf{P}\{\max_{1 \leq l \leq L} |\sum_{i=1}^n (z_{l,i} - \mathsf{E}_{i-1} z_{l,i})| > \kappa n^\upsilon\} = 0.$$

Lemma A.4. *(Johansen and Nielsen, 2016a, Theorem 5.3) For l so $1 \leq l \leq L$, let $z_{l,i}$ be $\mathcal{F}_i$ adapted satisfying $\mathsf{E} z_{l,i}^4 < \infty$. Let $D_q = \max_{1 \leq l \leq L} \sum_{i=1}^n \mathsf{E}_{i-1} z_{l,i}^{2^q}$ and suppose $\mathsf{E} D_q \leq Dn$ for $q = 1, 2$ and some $D > 0$. Then for all $\theta, \kappa > 0$*

$$\mathsf{P}\{\max_{1 \leq l \leq L} |\sum_{i=1}^n (z_{l,i} - \mathsf{E}_{i-1} z_{l,i})| > \kappa n^{1/2}\} \leq \frac{(L+1)\theta^3 D}{\kappa n} + \frac{\theta D}{\kappa} + 4L \exp(-\frac{\kappa\theta}{14}).$$

B Proofs of the empirical process results

We first prove uniform convergence and tightness for the one-sided empirical processes, and then linearize the compensator term. Finally extend results to the processes considering absolute residuals.

B.1 Uniform convergence for empirical process

We first consider Theorem B.1 which allows variation in a, b but c fixed. Then use the argument in Johansen and Nielsen (2016b, Theorem 4.1) to extend the result to Theorem B.2 where we have variation in b, c but $a = 0$, also see Berenguer, Johansen, and Nielsen (2018, Theorem 3.1), while use Jiao and Nielsen (2017, Theorem 5) to prove Theorem B.3 showing convergence uniformly in a, c but $b = 0$. Finally, by combining Theorem B.2, B.3 the uniform convergence in a, b, c follows.

Theorem B.1. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia, iia, iiia, va) holds with $\nu = 1$ and $s \geq 1$ such that $2^{s-1} > 1/2 + (1/4 - \eta)(2 + d_x)$. Then for any $0 \leq \psi \leq 1$, $B > 0$ and as $n \to \infty$*

$$\sup_{|a|,|b| \leq n^{1/4-\eta}B} |\mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)| = \mathrm{o}_{\mathsf{P}}(1).$$

Proof of Theorem B.1. Without loss of generality let $\sigma = 1$, $\Sigma^{1/2} = I_{d_x}$. Denote $R_n(a, b, c_\psi) = \mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)$ so we want to show $\mathcal{R}_{\psi,n} = \mathrm{o}_{\mathsf{P}}(1)$ for any $0 \leq \psi \leq 1$ as $n \to \infty$ where $\mathcal{R}_{\psi,n} = \sup_{|a|,|b| \leq n^{1/4-\eta}B} |R_n(a, b, c_\psi)|$. Throughout denote $C > 0$ as usual a constant which may have different values in different expressions.

1. *Construct a, b-balls*. For $\delta, n > 0$, cover the set $|a|, |b| \leq n^{1/4-\eta}B$ with balls of radius δ and centers a_j, b_m. The number of a, b-balls are $J = n^{1/4-\eta}B/\delta = \mathrm{O}(n^{1/4-\eta}/\delta)$ and $M = (n^{1/4-\eta}B/\delta)^{d_x} = \mathrm{O}(n^{(1/4-\eta)d_x}/\delta)$ respectively. Thus for any a, b there exist a_j, b_m so that $|a - a_j| \leq \delta$ and $|b - b_m| \leq \delta$.

2. *Apply chaining*. Write $R_n(a, b, c_\psi) = R_n(a_j, b_m, c_\psi) + \{R_n(a, b, c_\psi) - R_n(a_j, b_m, c_\psi)\}$ where $R_n(a_j, b_m, c_\psi)$ is a discrete point term while $R_n(a, b, c_\psi) - R_n(a_j, b_m, c_\psi)$ is a local oscillation term. The triangle inequality gives $\mathcal{R}_{\psi,n} \leq \mathcal{R}_{\psi,n,1} + \mathcal{R}_{\psi,n,2}$ where

$$\mathcal{R}_{\psi,n,1} = \max_{1 \leq j \leq J, 1 \leq m \leq M} |R_n(a_j, b_m, c_\psi)|,$$

$$\mathcal{R}_{\psi,n,2} = \max_{1 \leq j \leq J, 1 \leq m \leq M} \sup_{|a-a_j| \leq \delta, |b-b_m| \leq \delta} |R_n(a, b, c_\psi) - R_n(a_j, b_m, c_\psi)|.$$

It then suffices to show $\mathcal{R}_{\psi,n,1}, \mathcal{R}_{\psi,n,2}$ are $\mathrm{o}_{\mathsf{P}}(1)$.

3. *The discrete point term $\mathcal{R}_{\psi,n,1}$ is $\mathrm{o}_{\mathsf{P}}(1)$*. Recall the notation $J_{i,p}$ in (A.1) then write $R_n(a_j, b_m, c_\psi) = n^{-1/2} \sum_{i=1}^n (z_{l,i} - \mathsf{E}_{i-1} z_{l,i})$ where

$$z_{l,i} = w_{in} J_{i,p}(c_\psi, c_\psi + n^{-1/2} a_j c_\psi + z_{in}' \Pi b_m + n^{-1/2} r_i' b_m).$$

Use Lemma A.3 with $\upsilon = 1/2$, index $l = (j, m)$ and $L = JM$ to prove $\mathcal{R}_{\psi,n,1} = \mathrm{o}_{\mathsf{P}}(1)$, and we need to verify its conditions.

The moment $\mathsf{E}z_{l,i}^{2^s} < \infty$ *for* $s \in \mathbb{N}$. Bounding the difference of indicator functions by unity and using independence of u_i and $\mathcal{F}_{i-1}$ gives $\mathsf{E}z_{l,i}^{2^s} \leq \mathsf{E}|w_{in}|^{2^s} u_i^{2^s p} = \mathsf{E}|w_{in}|^{2^s}\mathsf{E}u_i^{2^s p}$. This is finite by Assumption 4.1(ia, va).

The parameter λ. The set of indices l has the size $L = JM$. Since $J = \mathrm{O}(n^{1/4-\eta}/\delta)$ and $M = \mathrm{O}(n^{(1/4-\eta)d_x}/\delta)$ while δ is fixed then $L = \mathrm{O}(n^\lambda)$ where $\lambda = (1/4-\eta)(1+d_x)$.

The parameter ς. Consider $1 \leq q \leq s$. Notice we have

$$|1_{(u_i \leq c_\psi + n^{-1/2}a_j c_\psi + z_{in}'\Pi b_m + n^{-1/2}r_i' b_m)} - 1_{(u_i \leq c_\psi)}| \leq 1_{(\underline{c}_{i\psi jm,r_i} \leq u_i \leq \overline{c}_{i\psi jm,r_i})},$$

where we denote

$$\begin{aligned}\underline{c}_{i\psi jm,r_i} &= c_\psi - n^{-1/2}|a_j||c_\psi| - |z_{in}||\Pi||b_m| - n^{-1/2}|r_i||b_m|,\\ \overline{c}_{i\psi jm,r_i} &= c_\psi + n^{-1/2}|a_j||c_\psi| + |z_{in}||\Pi||b_m| + n^{-1/2}|r_i||b_m|.\end{aligned}$$

Let $\mathcal{E}_i = \mathsf{E}_{i-1}J_{i,p}^{2^q}(c_\psi, c_\psi + n^{-1/2}a_j c_\psi + z_{in}'\Pi b_m + n^{-1/2}r_i' b_m)$, then by applying the above inequality concerning the difference of indicator functions it follows

$$\mathcal{E}_i \leq \iint_{y\in\mathbb{R}, x\in\mathbb{R}^{d_x}} y^{2^q p} 1_{(\underline{c}_{i\psi jm,x} \leq y \leq \overline{c}_{i\psi jm,x})} \mathsf{f}_{u,r}(y,x)dy(dx).$$

Decompose the joint density into the conditional and marginal ones, then apply the mean value theorem to the inner integral to get

$$\begin{aligned}\mathcal{E}_i &\leq \int_{x\in\mathbb{R}^{d_x}} \{\int_{\underline{c}_{i\psi jm,x}}^{\overline{c}_{i\psi jm,x}} (1+y^{2^s p})\mathsf{f}_{u|r}(y|x)dy\}\mathsf{f}_r(x)(dx)\\ &= \int_{x\in\mathbb{R}^{d_x}} (\overline{c}_{i\psi jm,x} - \underline{c}_{i\psi jm,x})(1+\tilde{c}^{2^s p})\mathsf{f}_{u|r}(\tilde{c}|x)\mathsf{f}_r(x)(dx),\end{aligned}$$

for an intermediate point $\tilde{c}$ such that $\underline{c}_{i\psi jm,x} \leq \tilde{c} \leq \overline{c}_{i\psi jm,x}$. Insert $\underline{c}_{i\psi jm,x}, \overline{c}_{i\psi jm,x}$ to get

$$\mathcal{E}_i \leq 2 \sup_{y\in\mathbb{R}, x\in\mathbb{R}^{d_x}} (1+y^{2^s p})\mathsf{f}_{u|r}(y|x) \int_{x\in\mathbb{R}^{d_x}} n^{-1/2}(|a_j||c_\psi| + |n^{1/2}z_{in}||\Pi||b_m| + |x||b_m|)\mathsf{f}_r(x)(dx).$$

Note $|a_j|, |b_m| \leq n^{1/4-\eta}B$ for each $1 \leq j \leq J$, $1 \leq m \leq M$ while c_ψ, Π are fixed. Furthermore since Assumption 4.1$(iia, iiia)$ that $\int_{x\in\mathbb{R}^{d_x}} |x|\mathsf{f}_r(x)(dx) < \infty$ and that $\sup_{y\in\mathbb{R},x\in\mathbb{R}^{d_x}}(1+y^{2^s p})\mathsf{f}_{u|r}(y|x) < \infty$, then there exists $C > 0$ so uniformly in j, m

$$\mathcal{E}_i \leq Cn^{-1/4-\eta}\int_{x\in\mathbb{R}^{d_x}} (1 + |n^{1/2}z_{in}| + |x|)\mathsf{f}_r(x)(dx) \leq Cn^{-1/4-\eta}(1+|n^{1/2}z_{in}|).$$

Insert the bound $\mathcal{E}_i$ we get

$$D_q = \max_{1\leq l\leq L}\sum_{i=1}^n \mathsf{E}_{i-1}z_{l,i}^{2^q} = \max_{1\leq j\leq J, 1\leq m\leq M}\sum_{i=1}^n |w_{in}|^{2^q}\mathcal{E}_i \leq Cn^{-1/4-\eta}\sum_{i=1}^n |w_{in}|^{2^q}(1+|n^{1/2}z_{in}|).$$

Since $|w_{in}|^{2^q} \leq (1+|w_{in}|^{2^s})$ and $\mathsf{E}\sum_{i=1}^n(1+|w_{in}|^{2^s})(1+|n^{1/2}z_{in}|) = \mathrm{O}(n)$ by Assumption 4.1(va), then we have $\mathsf{E}D_q = \mathrm{O}(n^\varsigma)$ where $\varsigma = 3/4 - \eta$.

Condition (i) is that $\varsigma < 2\upsilon$. This holds since $\eta > 0$ so $\varsigma = 3/4 - \eta < 1 = 2\upsilon$.

Condition (ii) is that $\varsigma + \lambda < \upsilon 2^s$. We have

$$\varsigma + \lambda = 3/4 - \eta + (1/4 - \eta)(1 + d_x) = 1/2 + (1/4 - \eta)(2 + d_x).$$

We choose s so that $\varsigma + \lambda < \upsilon 2^s = 2^{s-1}$.

4. *The oscillation term $\mathcal{R}_{\psi,n,2}$ is* $\mathsf{o}_\mathsf{P}(1)$. The same argument as item 3 shows $\mathcal{R}_{\psi,n,2}$ is small in probability. ■

Theorem B.2. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1$(i, iia, iiia, iva, va)$ holds with $\nu = 1$ and $s \geq 2$ satisfying (4.4). Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0 \leq \psi \leq 1} \sup_{|b| \leq n^{1/4-\eta} B} |\mathbb{F}_{u,n}^{w,p}(0, b, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)| = \mathsf{o}_\mathsf{P}(1).$$

Proof of Theorem B.2. Set $a = 0$ and then apply the argument by Johansen and Nielsen (2016a, Theorem 4.1) and Berenguer, Johansen, and Nielsen (2018, Theorem 3.1) to extend pointwise convergence shown in Theorem B.1 to convergence uniformly in ψ. It further requires chaining for the c_ψ-axis, so for $\delta, n > 0$ use the metric d to partition the quantile axis as laid out in (A.6) with $K = \text{int}(H_s n^{1/2}/\delta)$ where $s \geq 2$ such that (4.4) $2^{s-1} > 1 + (1/4 - \eta)(1 + d_x)$ holds. Notice $s \geq 2$ satisfying $2^{s-1} > 1 + (1/4 - \eta)(1 + d_x)$ implies $s \geq 1$ so that $2^{s-1} > 1/2 + (1/4 - \eta)(2 + d_x)$, so we need the higher moment condition on s for proving uniform convergence. ■

Theorem B.3. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1$(ia, ib, v)$ holds with $\nu = 1$, $s = 2$. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4-\eta} B} |\mathbb{F}_{u,n}^{w,p}(a, 0, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)| = \mathsf{o}_\mathsf{P}(1).$$

Proof of Theorem B.3. For $\delta, n > 0$ use the metric d to partition the c_ψ-axis as laid out in (A.6) with $K = \text{int}(H_s n^{1/2}/\delta)$ where $s = 2$, then argue as the proof of Theorem 5 in Jiao and Nielsen (2017) to show convergence uniformly in ψ, a. ■

Proof of Theorem 4.1. The term of interest is $\mathcal{W} = \mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)$. Denote $c_{\psi^\dagger} = c_\psi(1 + n^{-1/2}a/\sigma)$. Notice that $\mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) = \mathbb{F}_{u,n}^{w,p}(0, b, c_{\psi^\dagger})$ so it follows that $\mathcal{W} = \mathbb{F}_{u,n}^{w,p}(0, b, c_{\psi^\dagger}) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)$. Add and subtract $\mathbb{F}_{u,n}^{w,p}(a, 0, c_\psi) = \mathbb{F}_{u,n}^{w,p}(0, 0, c_{\psi^\dagger})$ and apply the triangle inequality to get

$$|\mathcal{W}| \leq |\mathbb{F}_{u,n}^{w,p}(0, b, c_{\psi^\dagger}) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_{\psi^\dagger})| + |\mathbb{F}_{u,n}^{w,p}(a, 0, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)|.$$

Thus, the problem reduces to showing

$$\sup_{0 \leq \psi^\dagger \leq 1} \sup_{|b| \leq n^{1/4-\eta} B} |\mathbb{F}_{u,n}^{w,p}(0, b, c_{\psi^\dagger}) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_{\psi^\dagger})| = \mathsf{o}_\mathsf{P}(1), \tag{B.1}$$

$$\sup_{0 \leq \psi \leq 1} \sup_{|a| \leq n^{1/4-\eta} B} |\mathbb{F}_{u,n}^{w,p}(a, 0, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)| = \mathsf{o}_\mathsf{P}(1). \tag{B.2}$$

Then (B.1) was considered by Theorem B.2 using Assumption 4.1$(i, iia, iiia, iva, va)$ with $\nu = 1$ and $s \geq 2$ such that (4.4) holds. Further, (B.2) follows by Theorem B.3, which requires Assumption 4.1(ia, ib, v) with $\nu = 1$, $s = 2$. ■

While the above is to show $n^{1/2}$-convergence result, we then prove n-convergence for a particular type of empirical process with weight $n^{1/2}z_{in} = z_i$ and marks r_i. We first demonstrate uniform convergence for the indicator process, then weights and marks can be incorporated by Hölder inequality.

Theorem B.4. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1 holds with $\nu = 1$ and $s \geq 2$ satisfying (4.4). Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B}\frac{1}{n}\sum_{i=1}^{n}|1_{(u_i\leq\sigma c_\psi+n^{-1/2}ac_\psi+z_{in}'\Pi b+n^{-1/2}r_i'b)} - 1_{(u_i\leq\sigma c_\psi)}| = \mathsf{O}_\mathsf{P}(n^{-1/4-\eta}).$$

Proof of Theorem B.4. Denote $\mathcal{R}_n' = \sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B} R_n'(a,b,c_\psi)$, where $g_i^{a,b,c_\psi} = 1_{(u_i\leq\sigma c_\psi+n^{-1/2}ac_\psi+z_{in}'\Pi b+n^{-1/2}r_i'b)} - 1_{(u_i\leq\sigma c_\psi)}$ and $R_n'(a,b,c_\psi) = n^{-1}\sum_{i=1}^n |g_i^{a,b,c_\psi}|$, then we show $\mathcal{R}_n' = \mathsf{O}_\mathsf{P}(n^{-1/4-\eta})$ as $n\to\infty$. We construct the martingale $R_{n,1}'(a,b,c_\psi)$ by adding and subtracting a compensator $R_{n,2}'(a,b,c_\psi)$ to $R_n'(a,b,c_\psi)$, therefore we have $R_n'(a,b,c_\psi) = R_{n,1}'(a,b,c_\psi) + R_{n,2}'(a,b,c_\psi)$ where

$$R_{n,1}'(a,b,c_\psi) = \frac{1}{n}\sum_{i=1}^{n}\{|g_i^{a,b,c_\psi}| - \mathsf{E}_{i-1}|g_i^{a,b,c_\psi}|\}, \qquad R_{n,2}'(a,b,c_\psi) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}_{i-1}|g_i^{a,b,c_\psi}|.$$

Triangle inequality gives $\mathcal{R}_n' \leq \mathcal{R}_{n,1}' + \mathcal{R}_{n,2}'$ where

$$\mathcal{R}_{n,1}' = \sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B}|R_{n,1}'(a,b,c_\psi)|, \qquad \mathcal{R}_{n,2}' = \sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B}R_{n,2}'(a,b,c_\psi).$$

It then suffices to prove $\mathcal{R}_{n,1}' = \mathsf{o}_\mathsf{P}(n^{-1/2})$ and $\mathcal{R}_{n,2}' = \mathsf{O}_\mathsf{P}(n^{-1/4-\eta})$. Set $w_{in} = 1$, $p = 0$ and adjust the rate n^{-1} instead of $n^{-1/2}$ to the empirical process, then use the argument in Theorem 4.1 to show $\mathcal{R}_{n,1}' = \mathsf{o}_\mathsf{P}(n^{-1/2})$, while $\mathcal{R}_{n,2}' = \mathsf{O}_\mathsf{P}(n^{-1/4-\eta})$ can be demonstrated by the reasoning of Theorem 4.2. ■

Proof of Theorem 4.4. Let $g_i^{a,b,c_\psi} = 1_{(u_i\leq\sigma c_\psi+n^{-1/2}ac_\psi+z_{in}'\Pi b+n^{-1/2}r_i'b)} - 1_{(u_i\leq\sigma c_\psi)}$ and $R_n^{z,r\prime}(a,b,c_\psi) = n^{-1/2}\sum_{i=1}^n z_{in}r_i'g_i^{a,b,c_\psi}$, then we want to show $\mathcal{R}_n^{z,r\prime} = \mathsf{o}_\mathsf{P}(1)$ as $n\to\infty$ where $\mathcal{R}_n^{z,r\prime} = \sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B}|R_n^{z,r\prime}(a,b,c_\psi)|$. By the triangle inequality we have $|R_n^{z,r\prime}(a,b,c_\psi)| \leq n^{-1}\sum_{i=1}^n |n^{1/2}z_{in}||r_i||g_i^{a,b,c_\psi}|$. Then apply Hölder inequality to get

$$|R_n^{z,r\prime}(a,b,c_\psi)| \leq (n^{-1}\sum_{i=1}^{n}|n^{1/2}z_{in}|^2|r_i|^2)^{1/2}(n^{-1}\sum_{i=1}^{n}|g_i^{a,b,c_\psi}|^2)^{1/2}.$$

Notice $|g_i^{a,b,c_\psi}|^2 = |g_i^{a,b,c_\psi}|$ and let $\mathcal{R}_n' = \sup_{0\leq\psi\leq 1}\sup_{|a|,|b|\leq n^{1/4-\eta}B} R_n'(a,b,c_\psi)$ where $R_n'(a,b,c_\psi) = n^{-1}\sum_{i=1}^n|g_i^{a,b,c_\psi}|$ so $\mathcal{R}_n^{z,r\prime} \leq (n^{-1}\sum_{i=1}^n |n^{1/2}z_{in}|^2|r_i|^2)^{1/2}(\mathcal{R}_n')^{1/2}$. Theorem B.4 shows $\mathcal{R}_n' = \mathsf{O}_\mathsf{P}(n^{-1/4-\eta})$ so $(\mathcal{R}_n')^{1/2} = \mathsf{O}_\mathsf{P}(n^{-1/8-\eta/2}) = \mathsf{o}_\mathsf{P}(1)$. Then it suffices to demonstrate $n^{-1}\sum_{i=1}^n |n^{1/2}z_{in}|^2|r_i|^2 = \mathsf{O}_\mathsf{P}(1)$. This is shown by Markov inequality using independence of z_i and r_i and Assumption 4.1(iia, va) that $\mathsf{E}|n^{1/2}z_{in}|^2, \mathsf{E}|r_i|^2 < \infty$. ■

B.2 Bias correction for compensator

Theorem 4.2 is proved by further considering two separate cases where variation in a, c_ψ but $b = 0$ and where in b, c_ψ but $a = 0$. The next theorem shows the bias correction term in the first situation where b is set to 0 and only a, c_ψ vary. The proof has the similar spirit as linearization of compensator in Jiao and Nielsen (2017, Theorem 8).

Theorem B.5. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia, ib, vb) holds with $\nu = 1$, $s = 0$. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\le\psi\le 1}\sup_{|a|\le n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{F}}^{w,p}_{u,n}(a,0,c_\psi) - \overline{\mathsf{F}}^{w,p}_{u,n}(0,0,c_\psi)\} - \mathcal{B}_{\mathsf{F}_u,n}(a,0,c_\psi)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}),$$

where the bias term is defined as

$$\mathcal{B}_{\mathsf{F}_u,n}(a,0,c_\psi) = \sigma^{p-1} c_\psi^p \mathsf{f}_u(c_\psi) n^{-1/2} \sum_{i=1}^n w_{in} n^{-1/2} a c_\psi.$$

Proof of Theorem B.5. See the proof of Theorem 8 in Jiao and Nielsen (2017) which is the generalized version of the above theorem, so follow their argument but set $b = 0$ in their context. ∎

We state next theorem to handle variation in b, c_ψ while a is set to 0.

Theorem B.6. *Let $c_\psi = \mathsf{F}_u^{-1}(\psi)$. Suppose Assumption 4.1(ia, iia, iiia, vb) holds with $\nu = 1$, $s = 0$. Then for any $B > 0$ and as $n \to \infty$*

$$\sup_{0\le\psi\le 1}\sup_{|b|\le n^{1/4-\eta}B} |n^{1/2}\{\overline{\mathsf{F}}^{w,p}_{u,n}(0,b,c_\psi) - \overline{\mathsf{F}}^{w,p}_{u,n}(0,0,c_\psi)\} - \mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi)| = \mathrm{O}_{\mathsf{P}}(n^{-2\eta}),$$

where the bias term, with $\xi_{c_\psi} = \mathsf{E}(\Sigma^{-1/2} r_i | u_i/\sigma = c_\psi)$, is defined as

$$\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi) = \sigma^{p-1} c_\psi^p \mathsf{f}_u(c_\psi) n^{-1/2} \sum_{i=1}^n w_{in}(n^{-1/2}\xi'_{c_\psi}\Sigma^{1/2} b + z'_{in}\Pi b).$$

Proof of Theorem B.6. The object of interest is

$$D_n(0,b,c_\psi) = n^{1/2}\{\overline{\mathsf{F}}^{w,p}_{u,n}(0,b,c_\psi) - \overline{\mathsf{F}}^{w,p}_{u,n}(0,0,c_\psi)\} - \mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi),$$

where $\overline{\mathsf{F}}^{w,p}_{u,n}(0,b,c_\psi)$ is well-defined due to Assumption 4.1(ia) as the indicator function in expectation is bounded by 1. Let $g_i^{0,b,c_\psi} = 1_{(u_i \le \sigma c_\psi + z'_{in}\Pi b + n^{-1/2} r'_i b)} - 1_{(u_i \le \sigma c_\psi)}$. Then denote $h_i(0,b,c_\psi) = (n^{-1/2}\xi'_{c_\psi}\Sigma^{1/2} b + z'_{in}\Pi b)/\sigma$ and $s(c_\psi) = c_\psi^p \mathsf{f}_u(c_\psi)$. We define that $S_i(0,b,c_\psi) = \mathsf{E}_{i-1} u_i^p g_i^{0,b,c_\psi} - \sigma^p s(c_\psi) h_i(0,b,c_\psi)$ such that $D_n(0,b,c_\psi)$ is expressed as $n^{-1/2}\sum_{i=1}^n w_{in} S_i(0,b,c_\psi)$. To apply Lemma A.2, let $Y = u_i/\sigma$ and $X = \Sigma^{-1/2} r_i$ so their joint density is $\mathsf{f}_{u,r}(y,x)$ for $y \in \mathbb{R}$ and $x \in \mathbb{R}^{d_x}$ while Y, X are independent of $\mathcal{F}_{i-1}$ and $z_i \in \mathcal{F}_{i-1}$. Also let $b_1 = z'_{in}\Pi b/\sigma \in \mathbb{R}$, $b_2 = n^{-1/2}\Sigma^{1/2} b \in \mathbb{R}^{d_x}$, and $c = c_\psi$.

By Assumption 4.1(iii) that the joint density can be decomposed into conditional and marginal ones, the lemma gives

$$|S_i(0,b,c_\psi)| \le \sigma^p\{\frac{1}{2}(z_{in}'\Pi b/\sigma)^2 + |z_{in}'\Pi b/\sigma||n^{-1/2}\Sigma^{1/2}b|\mathsf{E}|\Sigma^{-1/2}r_i| + \frac{1}{2}|n^{-1/2}\Sigma^{1/2}b|^2\mathsf{E}|\Sigma^{-1/2}r_i|^2\} \sup_{y\in\mathbb{R},x\in\mathbb{R}^{d_x}} |py^{p-1}\mathsf{f}_{u|r}(y|x) + y^p\dot{\mathsf{f}}_{u|r,y}(y|x)|,$$

uniformly in ψ. Note $|z_{in}'\Pi b| \le |z_{in}||\Pi||b|$ and $|\Sigma^{1/2}b| \le |\Sigma^{1/2}||b|$ due to the consistency property of the spectral norm. As $|b| \le n^{1/4-\eta}B$ and Assumption 4.1($iia, iiia$) with $s=0$ that $\mathsf{E}|\Sigma^{-1/2}r_i|, \mathsf{E}|\Sigma^{-1/2}r_i|^2 < \infty$ and $\sup_{y\in\mathbb{R},x\in\mathbb{R}^{d_x}} |py^{p-1}\mathsf{f}_{u|r}(y|x)+y^p\dot{\mathsf{f}}_{u|r,y}(y|x)| < \infty$, we have $|S_i(0,b,c_\psi)| = \mathrm{O}(n^{-1/2-2\eta})(1+|n^{1/2}z_{in}|^2)$. Then the triangle inequality gives

$$|D_n(0,b,c_\psi)| \le n^{-1/2}\sum_{i=1}^n |w_{in}||S_i(0,b,c_\psi)| = \mathrm{O}(n^{-2\eta})n^{-1}\sum_{i=1}^n |w_{in}|(1+|n^{1/2}z_{in}|^2).$$

By Assumption 4.1(vb), the term $|D_n(0,b,c_\psi)|$ has order $\mathrm{O}_\mathsf{P}(n^{-2\eta})$ uniformly in ψ, b. ■

Then we combine terms $\mathcal{B}_{\mathsf{F}_u,n}(a,0,c_\psi)$ and $\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi)$ to correct the bias in the case where variation in a,b,c_ψ is jointly considered.

Proof of Theorem 4.2. The interest is

$$D_n(a,b,c_\psi) = n^{1/2}\{\overline{\mathsf{F}}_{u,n}^{w,p}(a,b,c_\psi) - \overline{\mathsf{F}}_{u,n}^{w,p}(0,0,c_\psi)\} - \mathcal{B}_{\mathsf{F}_u,n}(a,b,c_\psi),$$

where $\overline{\mathsf{F}}_{u,n}^{w,p}(a,b,c_\psi)$ is well-defined by Assumption 4.1(ia). Let $c_{\psi^\dagger} = c_\psi(1+n^{-1/2}a/\sigma)$. Note $\overline{\mathsf{F}}_{u,n}^{w,p}(a,b,c_\psi) = \overline{\mathsf{F}}_{u,n}^{w,p}(0,b,c_{\psi^\dagger})$ and $\mathcal{B}_{\mathsf{F}_u,n}(a,b,c_\psi) = \mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi) + \mathcal{B}_{\mathsf{F}_u,n}(a,0,c_\psi)$. Add and subtract $n^{1/2}\overline{\mathsf{F}}_{u,n}^{w,p}(a,0,c_\psi) = n^{1/2}\overline{\mathsf{F}}_{u,n}^{w,p}(0,0,c_{\psi^\dagger})$ and $\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_{\psi^\dagger})$ and apply the trianlge inequality to obtain

$$|D_n(a,b,c_\psi)| \le |D_n(0,b,c_{\psi^\dagger})| + |D_n(a,0,c_\psi)| + |\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_{\psi^\dagger}) - \mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi)|.$$

Since c_ψ varies in $\mathbb{R}$ and $|a| \le n^{1/4-\eta}B$, then $c_{\psi^\dagger}$ also moves on the whole real line so variation in ψ, a transfers to variation in $\psi^\dagger$. Furthermore $c_{\psi^\dagger}$ converges to c_ψ as $n\to\infty$. Theorem B.6 shows $|D_n(0,b,c_{\psi^\dagger})| = \mathrm{O}_\mathsf{P}(n^{-2\eta})$ uniformly in $\psi^\dagger, b$ by Assumption 4.1($ia, iia, iiia, vb$). While due to Assumption 4.1(ia, ib, vb) Theorem B.5 demonstrates $|D_n(a,0,c_\psi)| = \mathrm{O}_\mathsf{P}(n^{-2\eta})$ uniformly in ψ, a. Notice

$$\xi_y = \mathsf{E}(\Sigma^{-1/2}r_i|\frac{u_i}{\sigma} = y) = \int_{\mathbb{R}^{d_x}} x\mathsf{f}_{r|u}(x|y)(dx),$$

then ξ_y is continuous uniformly in y due to the property of integral and Assumption 4.1(iii) that the conditional density $\mathsf{f}_{r|u}(x|y)$ is continuous uniformly in $x\in\mathbb{R}^{d_x}, y\in\mathbb{R}$. Also note y^p, $\mathsf{f}_u(y)$ are uniformly continuous in $y\in\mathbb{R}$ by Assumption 4.1(i). Since $c_{\psi^\dagger}$ only involves in $\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_{\psi^\dagger})$ through $c_{\psi^\dagger}^p$, $\mathsf{f}_u(c_{\psi^\dagger})$, $\xi_{c_{\psi^\dagger}}$, then $\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_{\psi^\dagger})$ converges to $\mathcal{B}_{\mathsf{F}_u,n}(0,b,c_\psi)$ uniformly in all its arguments as $c_{\psi^\dagger}\to c_\psi$, and subsequently the third term in the inequality of $|D_n(a,b,c_\psi)|$ vanishes. ■

B.3 Tightness for empirical process

The proof has the similar sprit as the asymptotic equi-continuity argument in Johansen and Nielsen (2016a, Theorem 4.4).

Proof of Theorem 4.3. Let $R(c_\psi, c_{\psi^\dagger}) = \mathbb{F}^{w,p}_{u,n}(0,0,c_{\psi^\dagger}) - \mathbb{F}^{w,p}_{u,n}(0,0,c_\psi)$. The aim is to bound the probability $\mathcal{P} = \mathsf{P}\{\mathcal{R} > \epsilon\}$ where $\mathcal{R} = \sup_{0\le\psi\le\psi^\dagger\le 1:\psi^\dagger-\psi\le\phi} |R(c_\psi, c_{\psi^\dagger})|$. Throughout, denote $C > 0$ as usual a constant not depending on ϵ, n, ϕ, which may have different values in different expressions. Assume $\sigma = 1$ without loss of generality and use the metric d built on the distance function $\mathsf{H}_s(x)$.

1. *Coefficients ϵ, ϕ, s.* Let $s = 2$. Note $0 < \epsilon < 1$ and $0 < \phi < 1$. Take ϵ, n as given and choose ϕ such that $\phi^{(1-\nu)/4} \le \epsilon^2$ for some $0 < \nu < 1$. A dyadic argument will be used, so given ϵ, n, ϕ we will select numbers $\overline{m}$, $\underline{m}$ and derive a bound to the probability $\mathcal{P}$ not depending on $\overline{m}$, $\underline{m}$.

2. *Fine grid.* Choose $\overline{m}$ so $2^{-\overline{m}}H_s \le n^{-1/2}\epsilon\phi^{(1-\nu)/4} \le 2^{1-\overline{m}}H_s$ where $H_s < \infty$ due to Assumption 4.1(ia).

3. *Coarse grid.* Choose $\underline{m}$ so $2^{-\underline{m}-1}H_s \le C\phi^{1-\nu} \le 2^{-\underline{m}}H_s$. For large n, $\overline{m} > \underline{m}$.

4. *Partition the support.* For each of $m = \underline{m}, \dots, \overline{m}$, use the metric d to partition $\mathbb{R}$ as laid out in (A.6) with $K_m = 2^m$, since $H_s < \infty$ by Assumption 4.1(ia).

5. *Assign c_ψ and $c_{\psi^\dagger}$ to the partitioned support.* For each of $m = \underline{m}, \dots, \overline{m}$, there exist $k_m \le k_m^\dagger$ and grid points $c_{k_m-1,m}$, $c_{k_m,m}$ and $c_{k_m^\dagger-1,m}$, $c_{k_m^\dagger,m}$ so $c_{k_m-1,m} < c_\psi \le c_{k_m,m}$ and $c_{k_m^\dagger-1,m} < c_{\psi^\dagger} \le c_{k_m^\dagger,m}$. Let $\underline{c}_m = c_{k_m-1,m}$, $\overline{c}_m = c_{k_m,m}$ and $\underline{c}_m^\dagger = c_{k_m^\dagger-1,m}$, $\overline{c}_m^\dagger = c_{k_m^\dagger,m}$ so $\underline{c}_m < c_\psi \le \overline{c}_m$ and $\underline{c}_m^\dagger < c_{\psi^\dagger} \le \overline{c}_m^\dagger$. Then $\overline{c}_{m-1} = c_{k_{m-1},m-1}$ equals either $\overline{c}_m = c_{k_m,m}$ or $c_{k_m+1,m}$ so $\mathsf{d}(\overline{c}_m, \overline{c}_{m-1})$ is either 0 or $2^{-m}H_s$. Since the bound $C\phi^{1-\nu} \le 2^{-\underline{m}}H_s$ in item 3 and Lemma A.1 using Assumption 4.1(ia) with some $0 < \nu < 1$, there exist $C_\nu, \phi_0 > 0$ so for $0 \le \phi \le \phi_0$

$$\sup_{0\le\psi\le\psi^\dagger\le 1:\psi^\dagger-\psi\le\phi} \mathsf{d}(c_\psi, c_{\psi^\dagger}) \le C_\nu\phi^{1-\nu} \le 2^{-\underline{m}}H_s.$$

Thus there is at most one $\underline{m}$-grid point in the interval $(c_\psi, c_{\psi^\dagger})$.

6. *Apply chaining.* Relate c_ψ to the nearest right fine grid point $\overline{c}_{\overline{m}}$ and $c_{\psi^\dagger}$ to the nearest left fine grid point $\underline{c}_{\overline{m}}^\dagger$. Then split the interval $(c_\psi, c_{\psi^\dagger})$ into the three intervals $(c_\psi, \overline{c}_{\overline{m}})$, $(\overline{c}_{\overline{m}}, \underline{c}_{\overline{m}}^\dagger)$, $(\underline{c}_{\overline{m}}^\dagger, c_{\psi^\dagger})$. If c_ψ, $c_{\psi^\dagger}$ are in the neighbouring $\overline{m}$-interval then $\overline{c}_{\overline{m}} = \underline{c}_{\overline{m}}^\dagger$, and if they are in the same $\overline{m}$-interval then $\overline{c}_{\overline{m}} > \underline{c}_{\overline{m}}^\dagger$ so $\underline{c}_{\overline{m}} = \underline{c}_{\overline{m}}^\dagger$ and $\overline{c}_{\overline{m}} = \overline{c}_{\overline{m}}^\dagger$. Thus

$$R(c_\psi, c_{\psi^\dagger}) = R(c_\psi, \overline{c}_{\overline{m}}) + R(\underline{c}_{\overline{m}}^\dagger, c_{\psi^\dagger}) - 1_{(\overline{c}_{\overline{m}} > \underline{c}_{\overline{m}}^\dagger)} R(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}}) + 1_{(\overline{c}_{\overline{m}} < \underline{c}_{\overline{m}}^\dagger)} R(\overline{c}_{\overline{m}}, \underline{c}_{\overline{m}}^\dagger).$$

An iterative argument can be made for the fourth term. Since $\overline{c}_{\overline{m}} < \underline{c}_{\overline{m}}^\dagger$, the coarser $(\overline{m}-1)$-grid points satisfy $\overline{c}_{\overline{m}} \le \overline{c}_{\overline{m}-1} \le \underline{c}_{\overline{m}-1}^\dagger \le \underline{c}_{\overline{m}}^\dagger$ so that

$$R(\overline{c}_{\overline{m}}, \underline{c}_{\overline{m}}^\dagger) = R(\overline{c}_{\overline{m}}, \overline{c}_{\overline{m}-1}) + R(\overline{c}_{\overline{m}-1}, \underline{c}_{\overline{m}-1}^\dagger) + R(\underline{c}_{\overline{m}-1}^\dagger, \underline{c}_{\overline{m}}^\dagger).$$

If $\overline{c}_{\overline{m}-1} = \underline{c}_{\overline{m}-1}^\dagger$, then $R(\overline{c}_{\overline{m}-1}, \underline{c}_{\overline{m}-1}^\dagger) = 0$ and iteration stops. In this case for $m < \overline{m}-1$ the m-grid points cross over as $\overline{c}_m \ge \overline{c}_{\overline{m}-1} = \underline{c}_{\overline{m}-1}^\dagger \ge \underline{c}_m^\dagger$. If $\overline{c}_{\overline{m}-1} < \underline{c}_{\overline{m}-1}^\dagger$, the argument can be made again for $R(\overline{c}_{\overline{m}-1}, \underline{c}_{\overline{m}-1}^\dagger)$. In the m-th step, iteration continues if $\overline{c}_m < \underline{c}_m^\dagger$, and noting that if there are no other m-grid points between $\overline{c}_m$, $\underline{c}_m^\dagger$, the contribution from the $(m-1)$-th step is zero. Item 5 shows there is at most one $\underline{m}$-grid point in the interval $(c_\psi, c_{\psi^\dagger})$, so the $\underline{m}$-th step either gives a zero contribution or the grid points have crossed over at an earlier stage. Therefore, the fourth term satisfies

$$1_{(\overline{c}_{\overline{m}}<\underline{c}^\dagger_{\overline{m}})}R(\overline{c}_{\overline{m}}, \underline{c}^\dagger_{\overline{m}}) = \sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m<\underline{c}^\dagger_m)}\{R(\overline{c}_m, \overline{c}_{m-1}) + R(\underline{c}^\dagger_{m-1}, \underline{c}^\dagger_m)\}.$$

Then apply the triangle inequality to get

$$|R(c_\psi, c_{\psi^\dagger})| \leq |R(c_\psi, \overline{c}_{\overline{m}})| + |R(\underline{c}^\dagger_{\overline{m}}, c_{\psi^\dagger})| + |1_{(\overline{c}_{\overline{m}}>\underline{c}^\dagger_{\overline{m}})}R(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})|$$
$$+ |\sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m<\underline{c}^\dagger_m)}R(\overline{c}_m, \overline{c}_{m-1})| + |\sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m<\underline{c}^\dagger_m)}R(\underline{c}^\dagger_{m-1}, \underline{c}^\dagger_m)|.$$

Thus it is argued that $\mathcal{R} \leq \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4 + \mathcal{R}_5$ where

$$\mathcal{R}_1 = \sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |R(c_\psi, \overline{c}_{\overline{m}})|,$$
$$\mathcal{R}_2 = \sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |R(\underline{c}^\dagger_{\overline{m}}, c_{\psi^\dagger})|,$$
$$\mathcal{R}_3 = \sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |1_{(\overline{c}_{\overline{m}}>\underline{c}^\dagger_{\overline{m}})}R(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})|,$$
$$\mathcal{R}_4 = \sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |\sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m<\underline{c}^\dagger_m)}R(\overline{c}_m, \overline{c}_{m-1})|,$$
$$\mathcal{R}_5 = \sup_{0\leq\psi\leq\psi^\dagger\leq 1:\psi^\dagger-\psi\leq\phi} |\sum_{m=\underline{m}+1}^{\overline{m}} 1_{(\overline{c}_m<\underline{c}^\dagger_m)}R(\underline{c}^\dagger_{m-1}, \underline{c}^\dagger_m)|.$$

By Boole's inequality, it follows that $\mathcal{P} \leq \mathcal{P}_1+\mathcal{P}_2+\mathcal{P}_3+\mathcal{P}_4+\mathcal{P}_5$ where $\mathcal{P}_1 = \mathsf{P}(\mathcal{R}_1 > \epsilon/5)$, $\mathcal{P}_2 = \mathsf{P}(\mathcal{R}_2 > \epsilon/5)$, $\mathcal{P}_3 = \mathsf{P}(\mathcal{R}_3 > \epsilon/5)$, $\mathcal{P}_4 = \mathsf{P}(\mathcal{R}_4 > \epsilon/5)$, $\mathcal{P}_5 = \mathsf{P}(\mathcal{R}_5 > \epsilon/5)$. It then suffices to find bounds to $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{P}_3$, $\mathcal{P}_4$, $\mathcal{P}_5$.

7. *Decompose the term* $\mathcal{R}_1$. Since $\underline{c}_{\overline{m}} < c_\psi \leq \overline{c}_{\overline{m}}$ and $\underline{c}_{\overline{m}} = c_{k_{\overline{m}}-1,\overline{m}}$, $\overline{c}_{\overline{m}} = c_{k_{\overline{m}},\overline{m}}$, we have $|J_{i,p}(c_\psi, \overline{c}_{\overline{m}})| \leq |J_{i,p}(\underline{c}_{\overline{m}}, \overline{c}_{\overline{m}})| = |J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})|$. Then by the triangle inequality, it follows

$$\mathcal{R}_1 \leq \max_{1\leq k_{\overline{m}}\leq K_{\overline{m}}} n^{-1/2}\sum_{i=1}^n |w_{in}|\{|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})| + \mathsf{E}_{i-1}|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})|\}.$$

Thus a martingale decomposition gives $\mathcal{R}_1 \leq \widetilde{\mathcal{R}}_1 + 2\overline{\mathcal{R}}_1$, where

$$\widetilde{\mathcal{R}}_1 = \max_{1\leq k_{\overline{m}}\leq K_{\overline{m}}} |n^{-1/2}\sum_{i=1}^n |w_{in}|\{|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})| - \mathsf{E}_{i-1}|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})|\}|,$$
$$\overline{\mathcal{R}}_1 = \max_{1\leq k_{\overline{m}}\leq K_{\overline{m}}} n^{-1/2}\sum_{i=1}^n |w_{in}|\mathsf{E}_{i-1}|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})|.$$

By Boole's inequality, it shows that $\mathcal{P}_1 \leq \widetilde{\mathcal{P}}_1 + 2\overline{\mathcal{P}}_1$ where we have $\widetilde{\mathcal{P}}_1 = \mathsf{P}(\widetilde{\mathcal{R}}_1 > \epsilon/15)$, $\overline{\mathcal{P}}_1 = \mathsf{P}(\overline{\mathcal{R}}_1 > \epsilon/15)$. It suffices to find bounds to $\widetilde{\mathcal{P}}_1$, $\overline{\mathcal{P}}_1$.

8. *Bounding the probability* $\widetilde{\mathcal{P}}_1$. Let $z_{l,i} = |w_{in}||J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})|$ and write $\widetilde{\mathcal{R}}_1$ as the maximum of $|n^{-1/2}\sum_{i=1}^n(z_{l,i} - \mathsf{E}_{i-1}z_{l,i})|$ over index $l = k_{\overline{m}}$ with $L = K_{\overline{m}} = 2^{\overline{m}}$. The difference of indicators can be bounded by one, so it holds for the moment condition that $\mathsf{E}z_{l,i}^4 \leq \mathsf{E}|w_{in}|^4 u_i^{4p} = \mathsf{E}|w_{in}|^4 \mathsf{E}u_i^{4p} < \infty$. This is because u_i is independent of $w_{in} \in \mathcal{F}_{i-1}$ and due to Assumption 4.1(ia, va) that $\mathsf{E}u_i^{4p}, \mathsf{E}|w_{in}|^4 < \infty$. Consider $1 \leq q \leq s$ and $s = 2$. The inequality of (A.5) shows uniformly in $l = k_{\overline{m}}$ that

$$\mathsf{E}_{i-1}z_{l,i}^{2q} = |w_{in}|^{2q}\mathsf{E}_{i-1}J_{i,p}^{2q}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}}) < (1 + |w_{in}|^{2^s})\mathsf{d}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}}).$$

As $\mathsf{d}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}}) = H_s/K_{\overline{m}}$ and Assumption 4.1(va) $n^{-1}\mathsf{E}\sum_{i=1}^n(1 + |w_{in}|^{2^s}) = \mathrm{O}(1)$, then we have $D = 2^{-\overline{m}}H_s$ such that

$$\mathsf{E}D_q = \mathsf{E}\max_{1\leq l\leq L}\sum_{i=1}^n \mathsf{E}_{i-1}z_{l,i}^{2q} < 2^{-\overline{m}}H_s\mathsf{E}\sum_{i=1}^n(1 + |w_{in}|^{2^s}) \leq Dn.$$

Apply Lemma A.4 with $\kappa = \epsilon/15$ to get

$$\widetilde{\mathcal{P}}_1 \leq \frac{15(2^{\overline{m}}+1)\theta^3 H_s}{2^{\overline{m}}\epsilon n} + \frac{15\theta H_s}{2^{\overline{m}}\epsilon} + 2^2 2^{\overline{m}}\exp(-\frac{\epsilon\theta}{210}) \leq \frac{C\theta^3}{\epsilon n} + \frac{C\theta}{2^{\overline{m}}\epsilon} + C2^{\overline{m}}\exp(-\frac{\epsilon\theta}{210}).$$

Choose $\theta = 210\epsilon^{-1}(\log 2^{2\overline{m}} + \log\phi^{-1})$. Use the inequality $(x+y)^3 \leq C(x^3+y^3)$ and then $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ and $n^{-1}\epsilon^2 \leq C2^{-2\overline{m}}\phi^{-(1-\nu)/2}$ implied by bounds in item 1, 2 to obtain

$$\widetilde{\mathcal{P}}_{1,1} = \frac{C\theta^3}{\epsilon n} \leq \frac{C}{\epsilon^4 n}(\overline{m}^3 + \log^3\frac{1}{\phi}) \leq C2^{-2\overline{m}}\phi^{-(1-\nu)/2}\phi^{-3(1-\nu)/4}(\overline{m}^3 + \log^3\frac{1}{\phi}).$$

Since $2^{-\overline{m}/2} \leq 2^{-\underline{m}/2} \leq C\phi^{(1-\nu)/2}$ implied by the bound in item 3, and the functions $m^3 2^{-m/2}$ and $\phi^{(1-\nu)/2}\log^3\phi^{-1}$ are bounded for $m \geq 1$ and $0 < \phi < 1$, we have

$$\widetilde{\mathcal{P}}_{1,1} \leq C(2^{-\overline{m}/2}\phi^{-(1-\nu)/2})^3(\overline{m}^3 2^{-\overline{m}/2} + 2^{-\overline{m}/2}\phi^{-(1-\nu)/2}\phi^{(1-\nu)/2}\log^3\frac{1}{\phi})\phi^{(1-\nu)/4} \leq C\phi^{(1-\nu)/4}.$$

By $\epsilon^{-2} \leq \phi^{-(1-\nu)/4}$ given in item 1, it holds

$$\widetilde{\mathcal{P}}_{1,2} = \frac{C\theta}{2^{\overline{m}}\epsilon} \leq \frac{C}{2^{\overline{m}}\epsilon^2}(\overline{m} + \log\frac{1}{\phi}) \leq C2^{-\overline{m}}\phi^{-(1-\nu)/4}(\overline{m} + \log\frac{1}{\phi}).$$

Rewrite this bound and argue as for $\widetilde{\mathcal{P}}_{1,1}$ to get

$$\widetilde{\mathcal{P}}_{1,2} \leq C2^{-\overline{m}/2}\phi^{-(1-\nu)/2}(\overline{m}2^{-\overline{m}/2} + 2^{-\overline{m}/2}\phi^{-(1-\nu)/2}\phi^{(1-\nu)/2}\log\frac{1}{\phi})\phi^{(1-\nu)/4} \leq C\phi^{(1-\nu)/4}.$$

For $0 < \phi < 1$ and $0 < \nu < 1$ it holds $\phi \leq \phi^{(1-\nu)/4}$ so that

$$\widetilde{\mathcal{P}}_{1,3} = C2^{\overline{m}}\exp(-\frac{\epsilon\theta}{210}) = C2^{-\overline{m}}\phi \leq C\phi^{(1-\nu)/4}.$$

In summary, $\widetilde{\mathcal{P}}_1 \leq \widetilde{\mathcal{P}}_{1,1} + \widetilde{\mathcal{P}}_{1,2} + \widetilde{\mathcal{P}}_{1,3} \leq C\phi^{(1-\nu)/4}$.

9. *Bounding the probability* $\overline{\mathcal{P}}_1$. The inequality of (A.5) shows that uniformly in $k_{\overline{m}}$ we have $\mathsf{E}_{i-1}|J_{i,p}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}})| < \mathsf{d}(c_{k_{\overline{m}}-1,\overline{m}}, c_{k_{\overline{m}},\overline{m}}) = H_s/K_{\overline{m}}$. Then by $K_{\overline{m}} = 2^{\overline{m}}$ and Assumption 4.1(va) it follows $\mathsf{E}\overline{\mathcal{R}}_1 < n^{-1/2}\mathsf{E}\sum_{i=1}^n |w_{in}|H_s/K_{\overline{m}} \leq Cn^{1/2}2^{-\overline{m}}H_s$. By Markov inequality and the bound $2^{-\overline{m}}H_s \leq n^{-1/2}\epsilon\phi^{(1-\nu)/4}$ in item 2, we get

$$\overline{\mathcal{P}}_1 \leq \frac{15\mathsf{E}\overline{\mathcal{R}}_1}{\epsilon} \leq \frac{Cn^{1/2}2^{-\overline{m}}H_s}{\epsilon} \leq C\phi^{(1-\nu)/4}.$$

10. *Bounding the probability* $\mathcal{P}_1$. Combine results in item $7-9$ to get $\mathcal{P}_1 \leq C\phi^{(1-\nu)/4}$.

11. *Bounding the probability* $\mathcal{P}_2$. This is similar as to show $\mathcal{P}_1 \leq C\phi^{(1-\nu)/4}$. Thus the same argument can be made to demonstrate $\mathcal{P}_2 \leq C\phi^{(1-\nu)/4}$ through item $7-10$.

12. *Bounding the probability* $\mathcal{P}_3$. Notice $\mathcal{R}_3 \leq \mathcal{R}_1$ so $\mathcal{P}_3 \leq \mathcal{P}_1 \leq C\phi^{(1-\nu)/4}$.

13. *Decompose the term* $\mathcal{R}_4$. From partition in item 4, 5, $\overline{c}_{m-1} = c_{k_{m-1},m-1}$ equals either $\overline{c}_m = c_{k_m,m}$ or $c_{k_m+1,m}$, then $\overline{c}_m$, $\overline{c}_{m-1}$ are at most 1 step apart in the m-grid so that $R(\overline{c}_m, \overline{c}_{m-1})$ is either 0 or $R(c_{k_m,m}, c_{k_m+1,m})$. Then we get

$$|R(\overline{c}_m, \overline{c}_{m-1})| \leq |R(c_{k_m,m}, c_{k_m+1,m})| \leq \max_{1\leq k_m \leq K_m} |R(c_{k_m-1,m}, c_{k_m,m})|.$$

Note that the right-hand side on the last inequality does not depend on ψ, $\psi^\dagger$ so

$$\mathcal{R}_4 \leq \sum_{m=\underline{m}+1}^{\overline{m}} \max_{1\leq k_m \leq K_m} |R(c_{k_m-1,m}, c_{k_m,m})|.$$

14. *Bounding the probability* $\mathcal{P}_4$. Notice first that

$$\sum_{m=\underline{m}+1}^{\overline{m}} 2^{(\underline{m}-m)/4} < \sum_{j=1}^{\infty} 2^{-j/4} = \frac{1}{2^{1/4}-1} < 6,$$

so $\sum_{m=\underline{m}+1}^{\overline{m}} 2^{(\underline{m}-m)/4}\epsilon/30 < \epsilon/5$. Then $\mathcal{R}_4 > \epsilon/5$ implies

$$\sum_{m=\underline{m}+1}^{\overline{m}} \max_{1\leq k_m \leq K_m} |R(c_{k_m-1,m}, c_{k_m,m})| > \sum_{m=\underline{m}+1}^{\overline{m}} \frac{2^{(\underline{m}-m)/4}\epsilon}{30}.$$

It therefore holds

$$\mathcal{P}_4 \leq \mathsf{P}[\bigcup_{m=\underline{m}+1}^{\overline{m}} \{\max_{1\leq k_m \leq K_m} |R(c_{k_m-1,m}, c_{k_m,m})| > \frac{2^{(\underline{m}-m)/4}\epsilon}{30}\}].$$

Then by Boole's inequality, we have

$$\mathcal{P}_4 \leq \sum_{m=\underline{m}+1}^{\overline{m}} \mathsf{P}\{\max_{1\leq k_m \leq K_m} |R(c_{k_m-1,m}, c_{k_m,m})| > \frac{2^{(\underline{m}-m)/4}\epsilon}{30}\}.$$

Let $z_{l,i} = w_{in}J_{i,p}(c_{k_m-1,m}, c_{k_m,m})$ and write $R(c_{k_m-1,m}, c_{k_m,m})$ as $n^{-1/2}\sum_{i=1}^n (z_{l,i} - \mathsf{E}_{i-1}z_{l,i})$, where l represents the index k_m with $L = K_m = 2^m$. The moment condition $\mathsf{E}z_{l,i}^4 < \infty$

holds due to Assumption 4.1(ia, va). Consider $1 \le q \le s$ and $s = 2$. The inequality of (A.5) shows uniformly in $l = k_m$

$$\mathsf{E}_{i-1} z_{l,i}^{2q} = |w_{in}|^{2q} \mathsf{E}_{i-1} J_{i,p}^{2q}(c_{k_m-1,m}, c_{k_m,m}) < (1 + |w_{in}|^{2^s}) \mathsf{d}(c_{k_m-1,m}, c_{k_m,m}).$$

Since $\mathsf{d}(c_{k_m-1,m}, c_{k_m,m}) = H_s/K_m$ and Assumption 4.1(va), we have $D = 2^{-m} H_s$ so that

$$\mathsf{E} D_q = \mathsf{E} \max_{1 \le l \le L} \sum_{i=1}^{n} \mathsf{E}_{i-1} z_{l,i}^{2q} < 2^{-m} H_s \mathsf{E} \sum_{i=1}^{n} (1 + |w_{in}|^{2^s}) \le Dn.$$

Apply Lemma A.4 with $\kappa = 2^{(\underline{m}-m)/4}\epsilon/30$ to get

$$\mathcal{P}_4 \le \sum_{m=\underline{m}+1}^{\overline{m}} \{ \frac{C\theta_m^3}{2^{(\underline{m}-m)/4}\epsilon n} + \frac{C\theta_m}{2^m 2^{(\underline{m}-m)/4}\epsilon} + C 2^m \exp(-\frac{2^{(\underline{m}-m)/4}\epsilon\theta_m}{420}) \}.$$

Choose $\theta_m = 420\epsilon^{-1} 2^{(m-\underline{m})/4}(\log 4^{m-\underline{m}} + \log \phi^{-1})$. For the first term above, use the inequality $(x+y)^3 \le C(x^3 + y^3)$ to get

$$\mathcal{P}_{4,1} = \sum_{m=\underline{m}+1}^{\overline{m}} \frac{C\theta_m^3}{2^{(\underline{m}-m)/4}\epsilon n} \le \sum_{m=\underline{m}+1}^{\overline{m}} \frac{C}{2^{\underline{m}-m}\epsilon^4 n} \{(m - \underline{m})^3 + \log^3 \frac{1}{\phi}\}.$$

Since $\epsilon^{-2} \le \phi^{-(1-\nu)/4}$ and $n^{-1}\epsilon^2 \le C 2^{-2\overline{m}} \phi^{-(1-\nu)/2}$ implied by bounds in item 1, 2, then

$$\mathcal{P}_{4,1} \le \sum_{m=\underline{m}+1}^{\overline{m}} C 2^{-2\overline{m}-\underline{m}+m} \phi^{-5(1-\nu)/4} \{(m - \underline{m})^3 + \log^3 \frac{1}{\phi}\}.$$

Rewrite $2^{-2\overline{m}-\underline{m}+m} = 2^{3(m-\overline{m})/2-(m-\underline{m})/2-\overline{m}/2-3\underline{m}/2}$ so

$$\mathcal{P}_{4,1} \le \sum_{m=\underline{m}+1}^{\overline{m}} C(2^{-\underline{m}/2}\phi^{-(1-\nu)/2})^3 \frac{2^{3(m-\overline{m})/2}}{2^{(m-\underline{m})/2}} \{ \frac{(m-\underline{m})^3}{2^{\overline{m}/2}} + \frac{\phi^{(1-\nu)/2}}{2^{\overline{m}/2}\phi^{(1-\nu)/2}} \log^3 \frac{1}{\phi} \} \phi^{(1-\nu)/4}.$$

Using $2^{-\overline{m}/2} \le 2^{-\underline{m}/2} \le C\phi^{(1-\nu)/2}$ from item 3 and that geometric sums are finite, argue as for $\widetilde{\mathcal{P}}_{1,1}$ in item 8 to obtain $\mathcal{P}_{4,1} \le C\phi^{(1-\nu)/4}$. Then the second term satisfies

$$\mathcal{P}_{4,2} = \sum_{m=\underline{m}+1}^{\overline{m}} \frac{C\theta_m}{2^m 2^{(\underline{m}-m)/4}\epsilon} \le \sum_{m=\underline{m}+1}^{\overline{m}} \frac{C}{2^{(\underline{m}+m)/2}\epsilon^2} \{(m - \underline{m}) + \log \frac{1}{\phi}\}.$$

Use $\epsilon^{-2} \le \phi^{-(1-\nu)/4}$ from item 1 to get

$$\mathcal{P}_{4,2} \le \sum_{m=\underline{m}+1}^{\overline{m}} C 2^{-(\underline{m}+m)/2} \phi^{-(1-\nu)/4} \{(m - \underline{m}) + \log \frac{1}{\phi}\}.$$

Rewrite $2^{-(\underline{m}+m)/2} = 2^{-(m-\underline{m})/2} 2^{-\underline{m}}$ so

$$\mathcal{P}_{4,2} \le \sum_{m=\underline{m}+1}^{\overline{m}} C 2^{-\underline{m}/2} \phi^{-(1-\nu)/2} 2^{-(m-\underline{m})/2} (\frac{m - \underline{m}}{2^{\underline{m}/2}} + \frac{\phi^{(1-\nu)/2}}{2^{\underline{m}/2}\phi^{(1-\nu)/2}} \log \frac{1}{\phi}) \phi^{(1-\nu)/4}.$$

Since $2^{-\underline{m}/2} \leq C\phi^{(1-\nu)/2}$ from item 3, argue as for $\mathcal{P}_{4,1}$ to show $\mathcal{P}_{4,2} \leq C\phi^{(1-\nu)/4}$. Then the third term satisfies

$$\mathcal{P}_{4,3} = \sum_{m=\underline{m}+1}^{\overline{m}} C2^m \exp(-\frac{2^{(\underline{m}-m)/4}\epsilon\theta_m}{420}) = \sum_{m=\underline{m}+1}^{\overline{m}} C2^{-(m-\underline{m})}2^{\underline{m}}\phi.$$

As $2^{\underline{m}} \leq C\phi^{\nu-1}$ implied by the bound in item 3, $\mathcal{P}_{4,3} \leq \sum_{m=\underline{m}+1}^{\overline{m}} C2^{-(m-\underline{m})}\phi^\nu \leq C\phi^\nu$. In summary, $\mathcal{P}_4 \leq \mathcal{P}_{4,1} + \mathcal{P}_{4,2} + \mathcal{P}_{4,3} \leq C\phi^{(1-\nu)/4} + C\phi^\nu$.

15. *Bounding the probability* $\mathcal{P}_5$. Apply the same argument as for $\mathcal{P}_4$ through item 13, 14 to show $\mathcal{P}_5 \leq C\phi^{(1-\nu)/4} + C\phi^\nu$.

16. *Tightness*. Combine the above result to obtain $\mathcal{P} \leq C\phi^{(1-\nu)/4} + C\phi^\nu$. For a given $0 < \epsilon < 1$ and a large n, the only constraint to $\phi \in (0,1)$ is $\phi^{(1-\nu)/4} \leq \epsilon^2$. Thus we have $\mathcal{P} \to 0$ as $\phi \downarrow 0$. ■

B.4 Results for empirical processes of absolute residuals

Empirical processes with absolute indicators can be expressed by the difference of two one-sided processes, thus we prove Theorem 4.5, 4.6, 4.7, 4.8 by using the one-sided process results in Theorem 4.1, 4.2, 4.3, 4.4. The absolute empirical process results are given under more restrictive Assumption 3.1, so we first investigate the lemma concerning the relationship between Assumption 3.1 and 4.1.

Lemma B.1. *Suppose* w_{in} *is either of* 1, $n^{1/2}z_{in} = z_i$, $nz_{in}z_{in}' = z_iz_i'$ *and* p *is either of* 0, 1, 2. *Then Assumption 3.1*($ia, iia, iiia, ivb, ivc$) *implies Assumption 4.1 with* $s \geq 2$ *satisfying (4.4).*

Proof of Theorem B.1. Assumption 3.1(ia) shows Assumption 4.1(ia, ic), while Assumption 4.1(ib) further needs continuous differentiability of the marginal density f_u, see discussion in Johansen and Nielsen (2016a, Remark 4.1c). Assumption 3.1(iia, ivb) is the same as Assumption 4.1(iia, iva), and Assumption 3.1($iiia$) deduces Assumption 4.1($iiia$) with continuous differentiability of the conditional density $\mathsf{f}_{u|r}$. At last Assumption 3.1(ivc) implies Assumption 4.1(va) and (vb) by Markov inequality. ■

Proof of Theorem 4.5. The term of interest is $\mathcal{G} = \mathbb{G}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{G}_{u,n}^{w,p}(0, 0, c_\psi)$. Our focus is on the absolute quantile $c_\psi = \mathsf{G}_u^{-1}(\psi) > 0$ rather than the one-sided quantile $c_{\psi^*} = \mathsf{F}_u^{-1}(\psi^*) \in \mathbb{R}$. Note $|u_i|/\sigma \sim \mathsf{G}_u$ and $u_i/\sigma \sim \mathsf{F}_u$. Since

$$\begin{aligned}&1_{(|u_i - z_{in}'\Pi b - n^{-1/2}r_i'b| \leq \sigma c_\psi + n^{-1/2}ac_\psi)}\\ =&1_{(u_i \leq \sigma c_\psi + n^{-1/2}ac_\psi + z_{in}'\Pi b + n^{-1/2}r_i'b)} - 1_{(u_i \leq -\sigma c_\psi - n^{-1/2}ac_\psi + z_{in}'\Pi b + n^{-1/2}r_i'b)}\end{aligned}$$

and by (4.3), (4.7), we have $\mathbb{G}_{u,n}^{w,p}(a, b, c_\psi) = \mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) - \lim_{c_\psi^\dagger \downarrow c_\psi} \mathbb{F}_{u,n}^{w,p}(a, b, -c_\psi^\dagger)$ for any $c_\psi > 0$. By this and the triangle inequality, then for any $c_\psi = \mathsf{G}_u^{-1}(\psi) > 0$,

$$|\mathcal{G}| \leq |\mathbb{F}_{u,n}^{w,p}(a, b, c_\psi) - \mathbb{F}_{u,n}^{w,p}(0, 0, c_\psi)| + \lim_{c_\psi^\dagger \downarrow c_\psi} |\mathbb{F}_{u,n}^{w,p}(a, b, -c_\psi^\dagger) - \mathbb{F}_{u,n}^{w,p}(0, 0, -c_\psi^\dagger)|.$$

These vanish uniformly in ψ, a, b by Theorem 4.1 using Assumption 4.1 with $\nu = 1$ and $s \geq 2$ such that (4.4) holds. Lemmma B.1 shows Assumption 3.1($ia, iia, iiia, ivb, ivc$) suffices. ■

Proof of Theorem 4.6. Argue as in the proof of Theorem 4.5 but using Theorem 4.2 instead of Theorem 4.1. Due to symmetry of f_u, approximate the compensator by

$$\mathcal{B}_{\mathsf{G}_u,n}(a,b,c_\psi) = \sigma^{p-1} c_\psi^p \mathsf{f}_u(c_\psi) n^{-1/2} \sum_{i=1}^n w_{in}[\{1+(-1)^p\} n^{-1/2} a c_\psi + n^{-1/2}\{\xi_{c_\psi} - (-1)^p \xi_{-c_\psi}\}' \Sigma^{1/2} b + \{1-(-1)^p\} z_{in}' \Pi b]. \qquad \blacksquare$$

Proof of Theorem 4.7. The same argument as Theorem 4.5 applies while the tightness result in Theorem 4.3 is used in this proof instead. ■

Proof of Theorem 4.8. Argue along the lines of the proof in Theorem 4.5, however replace Theorem 4.1 by Theorem 4.4. ■

C Proofs of the main results

We first present some preliminary results for stochastic expansions of two stage least squares statistics. Asymptotics are shown for Algorithm 2.1, then weak convergence for Algorithm 2.2 and 2.3.

C.1 Auxillary results for product moments

The robust estimators in Algorithm 2.1 are two stage least squares estimators for selected observations. Outliers are detected through the absolute value of estimated structural errors u_i adjusted by its scale σ in (2.1). Introduce the indicators in the empirical processes in §4 as

$$v_i^{a,b,c} = 1_{(|u_i - z_{in}' \Pi b - n^{-1/2} r_i' b| \le \sigma c + n^{-1/2} a c)}, \tag{C.1}$$

so given the cut-off c the indicators (2.9) for selecting non-outlying observations have the relationship

$$v_{i,c}^{(m)} = 1_{(|y_i - x_i' \widehat{\beta}_c^{(m)}| \le \widehat{\sigma}_c^{(m)} c)} = 1_{(|u_i - z_{in}' \Pi \widehat{b}_c^{(m)} - n^{-1/2} r_i' \widehat{b}_c^{(m)}| \le \sigma c + n^{-1/2} \widehat{a}_c^{(m)} c)} = v_i^{\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c}, \tag{C.2}$$

where $\widehat{b}_c^{(m)} = n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$, $\widehat{a}_c^{(m)} = n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are the estimation errors for β, σ. Thus in Algorithm 2.1 the least squares statistics in the estimators (2.10), (2.11), (2.12) for Π, β, σ are expressed by the weighted and marked empirical processes analyzed in section §4. The following results describe the asymptotic behaviour of the corresponding product moments.

Lemma C.1. *Suppose Assumption 3.1*$(ia, iia, iiia, ivb, ivc)$ *holds. Then we have*

$$n^{-1/2}\sum_{i=1}^{n} v_i^{a,b,c} = n^{-1/2}\sum_{i=1}^{n} 1_{(|u_i|\leq \sigma c)} + \mathsf{f}_u(c)\frac{2ac+\zeta_c^{-\prime}\Sigma^{1/2}b}{\sigma} + R_v(a,b,c),$$

$$n^{-1/2}\sum_{i=1}^{n} u_i^2 v_i^{a,b,c} = n^{-1/2}\sum_{i=1}^{n} u_i^2 1_{(|u_i|\leq \sigma c)} + \sigma^2 c^2 \mathsf{f}_u(c)\frac{2ac+\zeta_c^{-\prime}\Sigma^{1/2}b}{\sigma} + R_{vuu}(a,b,c),$$

$$\sum_{i=1}^{n} z_{in} u_i v_i^{a,b,c} = \sum_{i=1}^{n} z_{in} u_i 1_{(|u_i|\leq \sigma c)} + c\mathsf{f}_u(c)\sum_{i=1}^{n} z_{in}(n^{-1/2}\zeta_c^{+\prime}\Sigma^{1/2} + 2z_{in}'\Pi)b + R_{vzu}(a,b,c),$$

$$n^{1/2}\sum_{i=1}^{n} z_{in} z_{in}' v_i^{a,b,c} = n^{1/2}\sum_{i=1}^{n} z_{in} z_{in}' 1_{(|u_i|\leq \sigma c)} + \mathsf{f}_u(c) M_{zz,n}\frac{2ac+\zeta_c^{-\prime}\Sigma^{1/2}b}{\sigma} + R_{vzz}(a,b,c).$$

Let $R(a,b,c) = |R_v(a,b,c)| + |R_{vuu}(a,b,c)| + |R_{vzu}(a,b,c)| + |R_{vzz}(a,b,c)|$. *Then for any* $B > 0$ *and as* $n \to \infty$

$$\sup_{0<c<\infty}\ \sup_{|a|,|b|\leq n^{1/4-\eta}B} |R(a,b,c)| = \mathrm{o}_{\mathsf{P}}(1).$$

Proof of Lemma C.1. The terms of interest are

$$\mathcal{M}_n = n^{-1/2}\sum_{i=1}^{n} w_{in} u_i^p v_i^{a,b,c}, \quad v_i^{a,b,c} = 1_{(|u_i - z_{in}'\Pi b - n^{-1/2} r_i' b| \leq \sigma c + n^{-1/2} ac)}.$$

1. *Decompose* $\mathcal{M}_n$. Write $\mathcal{M}_n = \mathcal{M}_{n,1} + \mathcal{M}_{n,2} + \mathcal{M}_{n,3}$, where

$$\mathcal{M}_{n,1} = n^{-1/2}\sum_{i=1}^{n} w_{in} u_i^p 1_{(|u_i|\leq \sigma c)}, \quad \mathcal{M}_{n,2} = n^{-1/2}\sum_{i=1}^{n} w_{in}\mathsf{E}_{i-1} u_i^p\{v_i^{a,b,c} - 1_{(|u_i|\leq \sigma c)}\},$$

$$\mathcal{M}_{n,3} = n^{-1/2}\sum_{i=1}^{n} w_{in} u_i^p\{v_i^{a,b,c} - 1_{(|u_i|\leq \sigma c)}\} - n^{-1/2}\sum_{i=1}^{n} w_{in}\mathsf{E}_{i-1} u_i^p\{v_i^{a,b,c} - 1_{(|u_i|\leq \sigma c)}\}.$$

Therefore, the first term in stochastic expansion is $\mathcal{M}_{n,1}$. We will linearize $\mathcal{M}_{n,2}$ to obtain the second term, and argue that $\mathcal{M}_{n,3}$ is small in probability.

2. *Linearize* $\mathcal{M}_{n,2}$. Note $\mathcal{M}_{n,2} = n^{1/2}\{\overline{\mathsf{G}}_{u,n}^{w,p}(a,b,c) - \overline{\mathsf{G}}_{u,n}^{w,p}(0,0,c)\}$, see (4.6). Apply Theorem 4.6 by Assumption 3.1$(ia, iia, iiia, ivc)$ to get $\mathcal{M}_{n,2} = \mathcal{B}_{\mathsf{G}_{u},n}(a,b,c) + \mathrm{O}_{\mathsf{P}}(n^{-2\eta})$. Then $\mathcal{B}_{\mathsf{G}_{u},n}(a,b,c)$ reduces as desired. Note $0 < \eta \leq 1/4$ so $\mathcal{M}_{n,2} = \mathcal{B}_{\mathsf{G}_{u},n}(a,b,c) + \mathrm{o}_{\mathsf{P}}(1)$ uniformly in $0 < c < \infty$ and $|a|,|b| \leq n^{1/4-\eta}B$.

3. *Bounding* $\mathcal{M}_{n,3}$. Note $\mathcal{M}_{n,3} = \mathbb{G}_{u,n}^{w,p}(a,b,c) - \mathbb{G}_{u,n}^{w,p}(0,0,c)$, see (4.7). By Assumption 3.1$(ia, iia, iiia, ivb, ivc)$, Theorem 4.5 shows $\mathcal{M}_{n,3} = \mathrm{o}_{\mathsf{P}}(1)$ uniformly in a, b, c. ■

While the above results are asymptotic expansions with $n^{1/2}$ rate for two stage least squares statistics, stochastic approximation with rate n are also required to demonstrate the main asymptotic results for the estimators and gauge. Thus we establish some n-convergence expansions in the following.

Lemma C.2. *Suppose Assumption 3.1(ia, iia, iiia, iv) holds. Then we have*

$$n^{-1}\sum_{i=1}^{n} v_i^{a,b,c} = n^{-1}\sum_{i=1}^{n} 1_{(|u_i|\le\sigma c)} + R'_v(a,b,c),$$
$$\sum_{i=1}^{n} z_{in}z'_{in}v_i^{a,b,c} = \sum_{i=1}^{n} z_{in}z'_{in}1_{(|u_i|\le\sigma c)} + R'_{vzz}(a,b,c),$$

where for any $B>0$ and as $n\to\infty$

$$\sup_{0<c<\infty}\ \sup_{|a|,|b|\le n^{1/4-\eta}B} |R'_v(a,b,c)| + |R'_{vzz}(a,b,c)| = \mathrm{o_P}(1).$$

Proof of Lemma C.2. Adjust the first and fourth items in Lemma C.1 by the rate $n^{-1/2}$ and then observe the second terms in the expansions. We find $\sup_{c\in\mathbb{R}_+} c\mathsf{f}_u(c)<\infty$, $\sup_{c\in\mathbb{R}_+}|\zeta_c^-|\mathsf{f}_u(c)<\infty$ by Assumption 3.1(ia), and $n^{-1/2}|a|, n^{-1/2}|b|\le n^{-1/4-\eta}B$, while $M_{zz,n}\overset{\mathsf{P}}{\to} M_{zz}$ by Assumption 3.1(iva) and M_{zz}, Σ, σ are treated as constant. Notice that ζ_c^- can be expressed as the function of c, for instance under normality (2.5) it holds $\zeta_c^- = 2\Omega c$. Indeed the second terms in expansions are $\mathrm{O_P}(n^{-1/4-\eta})$ whereas the third terms are $\mathrm{o_P}(n^{-1/2})$ so combine them to show the remainder terms are $\mathrm{o_P}(1)$. ■

Lemma C.3. *Suppose Assumption 3.1(ia, iia, iiia, ivb, ivc) holds. Then we have*

$$n^{-1/2}\sum_{i=1}^{n} z_{in}r'_i v_i^{a,b,c} = n^{-1/2}\sum_{i=1}^{n} z_{in}r'_i 1_{(|u_i|\le\sigma c)} + R'_{vzr}(a,b,c),$$

where for any $B>0$ and as $n\to\infty$

$$\sup_{0<c<\infty}\ \sup_{|a|,|b|\le n^{1/4-\eta}B} |R'_{vzr}(a,b,c)| = \mathrm{o_P}(1).$$

Proof of Lemma C.3. Decompose the term of interest as

$$n^{-1/2}\sum_{i=1}^{n} z_{in}r'_i v_i^{a,b,c} = n^{-1/2}\sum_{i=1}^{n} z_{in}r'_i 1_{(|u_i|\le\sigma c)} + n^{-1/2}\sum_{i=1}^{n} z_{in}r'_i\{v_i^{a,b,c} - 1_{(|u_i|\le\sigma c)}\}.$$

Then Theorem 4.8 shows the second term in the above is $\mathrm{o_P}(1)$ uniformly in a,b,c. ■

C.2 Results for Algorithm 2.1

Since the indicators $v_{i,c}^{(m)}$ in 2SLS statistics equals to $v_i^{\widehat{a}_c^{(m)},\widehat{b}_c^{(m)},c}$, see (C.2), all processes $M_n(\widehat{a},\widehat{b},c)$ we considered depend on estimation errors $\widehat{b}=n^{1/2}(\widehat{\beta}-\beta)$, $\widehat{a}=n^{1/2}(\widehat{\sigma}-\sigma)$. If $\widehat{a},\widehat{b}$ are bounded by a compact set in probability, then $M_n(\widehat{a},\widehat{b},c)$ can be studied by analyzing the behaviour of $M_n(a,b,c)$ uniformly in $a,b\in\Theta$ for a compact set Θ. This is due to the lemma below.

Lemma C.4. *If for any $\epsilon>0$ there exists a compact set Θ so $\mathsf{P}(\widehat{a},\widehat{b}\in\Theta^c)<\epsilon$ as $n\to\infty$. Then we have $\mathsf{P}\{M_n(\widehat{a},\widehat{b},c)>\epsilon\}\le\mathsf{P}\{\sup_{a,b\in\Theta}|M_n(a,b,c)|>\epsilon\}+\epsilon$.*

Proof of Lemma C.4. Let $\mathcal{A} = \{|M_n(\widehat{a}, \widehat{b}, c)| > \epsilon\}, \mathcal{B} = (\widehat{a}, \widehat{b} \in \Theta)$. Then Boole's inequality shows $\mathsf{P}(\mathcal{A}) \leq \mathsf{P}(\mathcal{A} \cap \mathcal{B}) + \mathsf{P}(\mathcal{B}^c)$ as $\mathcal{A} = \mathcal{A} \cap (\mathcal{B} \cup \mathcal{B}^c) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{B}^c)$. The probability $\mathsf{P}(\mathcal{A} \cap \mathcal{B})$ is bounded by $\mathsf{P}\{\sup_{a,b \in \Theta} |M_n(a, b, c)| > \epsilon\}$ while $\mathsf{P}(\mathcal{B}^c) < \epsilon$ by assumption. ∎

First demonstrate consistency of $\widehat{\Pi}_c^{(m+1)}$ uniformly in $c \in [c_+, \infty)$ for a finite number $c_+ > 0$ given that $\widehat{b}_c^{(m)} = n^{1/2}(\widehat{\beta}_c^{(m)} - \beta)$, $\widehat{a}_c^{(m)} = n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma)$ are bounded in probability for any $m \in [0, \infty)$. The lemma below explores the uniform limiting behaviour of least squares statistics in the first stage regression (2.2).

Lemma C.5. *Let $c_\psi = \mathsf{G}_u^{-1}(\psi)$. Suppose Assumption 3.1(iia, ivc) holds. Then as $n \to \infty$*

$$\sup_{0 \leq \psi \leq 1} |n^{-1/2} \sum_{i=1}^{n} z_{in} r_i' 1_{(|u_i| \leq \sigma c_\psi)}| = \mathsf{o}_\mathsf{P}(1).$$

Proof of Lemma C.5. First observe that ψ is varying in the compact set $[0, 1]$. Moreover $z_i r_i' 1_{(|u_i| \leq \sigma c_\psi)}$ is continuous in $\psi \in [0, 1]$ with probability 1 since the only discontinuous point for c_ψ is at the realized value of u_i/σ while u_i/σ is assumed to have the continuous density f_u by Assumption 3.1. Then $|z_i r_i' 1_{(|u_i| \leq \sigma c_\psi)}| \leq |z_i||r_i|$ and $\mathsf{E}|z_i||r_i| < \infty$ since z_i is independent of r_i, u_i by Assumption 3.1 and $\mathsf{E}|z_i|, \mathsf{E}|r_i| < \infty$ by Assumption 3.1(iia, ivc). Therefore apply Uniform Law of Large Numbers, see Theorem 2 in Jennrich (1969), to show

$$n^{-1} \sum_{i=1}^{n} z_i r_i' 1_{(|u_i| \leq \sigma c_\psi)} \overset{\mathsf{P}}{\to} \mathsf{E} z_i r_i' 1_{(|u_i| \leq \sigma c_\psi)} = \mathsf{E} z_i \mathsf{E} r_i' 1_{(|u_i| \leq \sigma c_\psi)} = 0_{d_z \times d_x}$$

uniformly in ψ. The last equality follows $\mathsf{E} z_i = 0_{d_z}$ and $\mathsf{E}|r_i' 1_{(|u_i| \leq \sigma c_\psi)}| \leq \mathsf{E}|r_i| < \infty$. ∎

Remark C.1. *Notice in general that $\mathsf{E} r_i' 1_{(|u_i| \leq \sigma c_\psi)} \neq 0_{d_x}'$ except when $\psi = 1$ and $c_\psi = \infty$ so $\mathsf{E} r_i' 1_{(|u_i| \leq \sigma c_\psi)} = \mathsf{E} r_i' = 0_{d_x}'$. Write the expectation as an integral*

$$\mathsf{E} r_i' 1_{(|u_i| \leq \sigma c_\psi)} = \int_{x \in \mathbb{R}^{d_x}} \int_{-c_\psi}^{c_\psi} x' \mathsf{f}_{u,r}(y, x) dy (dx) \Sigma^{1/2}.$$

The above expected value equals to $0_{d_x}'$ if for any $y \in \mathbb{R}, x \in \mathbb{R}^{d_x}$ the density $\mathsf{f}_{u,r}$ satisfies $\mathsf{f}_{u,r}(y, x) = -\mathsf{f}_{u,r}(y, -x)$ or $\mathsf{f}_{u,r}(y, x) = -\mathsf{f}_{u,r}(-y, x)$ unless $\psi = 1$. However, symmetricity does not hold for the joint density in its second argument, for instance, under the normality assumption (2.5) when $\Omega \neq 0_{d_x}$, while obviously $\mathsf{f}_{u,r}$ is not an odd function.

Proof of Theorem 3.1. For any $m \in [0, \infty)$ the $m + 1$ step estimator for Π defined in (2.10) satisfies

$$\widehat{\Pi}_c^{(m+1)} - \Pi = (\sum_{i=1}^{n} z_{in} z_{in}' v_{i,c}^{(m)})^{-1} (n^{-1/2} \sum_{i=1}^{n} z_{in} r_i' v_{i,c}^{(m)}). \tag{C.3}$$

First notice $v_{i,c}^{(m)} = v_i^{\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c}$, see (C.2). Suppose $|\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathsf{O}_\mathsf{P}(1)$ and by Assumption 3.1($ia, iia, iiia, iv$), Lemma C.2 and Lemma C.3 show asymptotic expansions for least squares statistics for non-outlying observations. Lemma C.4 allows to replace deterministic terms b, a by stochastic estimation errors $\widehat{b}_c^{(m)}, \widehat{a}_c^{(m)}$. Substitute these expansions into (C.3) to obtain

$$\widehat{\Pi}_c^{(m+1)} - \Pi = (\sum_{i=1}^n z_{in}z_{in}' 1_{(|u_i|\le\sigma c)})^{-1}(n^{-1/2}\sum_{i=1}^n z_{in}r_i' 1_{(|u_i|\le\sigma c)}) + R_\Pi'(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where $R_\Pi'(a,b,c)$ vanishes uniformly in $c_+ \le c < \infty$ and $|a|, |b| \le B$. Check the moment condition $\mathsf{E}|z_i z_i' 1_{(|u_i|\le\sigma c)}| \le \mathsf{E}|z_i|^2 < \infty$ by Assumption 3.1(ivc), Law of Large Numbers shows for any $c \in [c_+, \infty)$

$$n^{-1}\sum_{i=1}^n z_i z_i' 1_{(|u_i|\le\sigma c)} \overset{\mathsf{P}}{\to} \mathsf{E}z_i z_i' 1_{(|u_i|\le\sigma c)} = \mathsf{E}z_i z_i' \mathsf{E}1_{(|u_i|\le\sigma c)} = M_{zz}\psi,$$

where the first equality is because z_i is independent of u_i. Theorem 4.7 demonstrates tightness of the above sequence so $\sum_{i=1}^n z_{in}z_{in}' 1_{(|u_i|\le\sigma c)}$ converges in probability to $M_{zz}\psi$ uniformly in $c \in [c_+, \infty)$. A key to this is that c is bounded away from 0 and $M_{zz} > 0$ by Assumption 3.1(iva) so $\psi, M_{zz}\psi > 0$. Thus further by Lemma C.5 due to Assumption 3.1(iia, ivc) and by Slutsky's theorem uniform consistency then follows. ■

Then we prove the theorem about stochastic expansion of the updated estimator in Algorithm 2.1.

Proof of Theorem 3.2. First notice that by substituting the Π estimator (2.10) the inverse term in the updated β estimator (2.11) is expressed as

$$\begin{aligned}\widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_i z_i' v_{i,c}^{(m)} \widehat{\Pi}_c^{(m+1)} &= \widehat{\Pi}_c^{(m+1)\prime}(\sum_{i=1}^n z_i z_i' v_{i,c}^{(m)})(\sum_{i=1}^n z_i z_i' v_{i,c}^{(m)})^{-1}(\sum_{i=1}^n z_i x_i' v_{i,c}^{(m)}) \\ &= \widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_i x_i' v_{i,c}^{(m)}. \end{aligned} \tag{C.4}$$

Apply (C.4) and (2.1) into the $m+1$ step β estimator (2.11) so for any $m \in [0, \infty)$

$$\widehat{\beta}_c^{(m+1)} = \beta + (\widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_i x_i' v_{i,c}^{(m)})^{-1}(\widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_i u_i v_{i,c}^{(m)}).$$

Use (C.4) again and adjust the above equation by the rate $n^{1/2}$ to get

$$n^{1/2}(\widehat{\beta}_c^{(m+1)} - \beta) = (\widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_{in} z_{in}' v_{i,c}^{(m)} \widehat{\Pi}_c^{(m+1)})^{-1}(\widehat{\Pi}_c^{(m+1)\prime}\sum_{i=1}^n z_{in} u_i v_{i,c}^{(m)}). \tag{C.5}$$

Suppose $|\widehat{b}_c^{(m)}| + |\widehat{a}_c^{(m)}| = \mathrm{O}_{\mathsf{P}}(1)$, Theorem 3.1 by Assumption 3.1($ia, iia, iiia, iv$) shows $\widehat{\Pi}_c^{(m+1)} \overset{\mathsf{P}}{\to} \Pi$ uniformly in $c \in [c_+, \infty)$. Apply the same argument in the proof of Theorem

3.1, so by Assumption 3.1$(ia, iia, iiia, iv)$ it allows to substitute expansions in Lemma C.1 and C.2 into (C.5). Then by Slutsky's theorem we have

$$\widehat{b}_c^{(m+1)} = (\sum_{i=1}^n \tilde{x}_{in}\tilde{x}_{in}' 1_{(|u_i|\leq\sigma c)})^{-1}\{c\mathsf{f}_u(c)\sum_{i=1}^n \tilde{x}_{in}(n^{-1/2}\zeta_c^{+\prime}\Sigma^{1/2} + 2\tilde{x}_{in}')\widehat{b}_c^{(m)} + \sum_{i=1}^n \tilde{x}_{in}u_i 1_{(|u_i|\leq\sigma c)}\} + R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where $R_\beta(a,b,c)$ vanishes uniformly in $c_+ \leq c < \infty$ and $|a|, |b| \leq B$. Explore the inverse term in the above equation

$$\sum_{i=1}^n \tilde{x}_{in}\tilde{x}_{in}' 1_{(|u_i|\leq\sigma c)} = \sum_{i=1}^n \tilde{x}_{in}\tilde{x}_{in}'\psi + \sum_{i=1}^n \tilde{x}_{in}\tilde{x}_{in}'(1_{(|u_i|\leq\sigma c)} - \psi) = M_{\tilde{x}\tilde{x},n}\psi + \mathrm{o}_{\mathsf{P}}(1),$$

where the second term converges in probability to zero uniformly in c by Law of Large Numbers and tightness property shown in Theorem 4.7, see the similar argument in the proof of Theorem 3.1. Note $n^{-1/2}\sum_{i=1}^n \tilde{x}_{in} \overset{\mathsf{P}}{\to} \mathsf{E}\tilde{x}_i = \Pi'\mathsf{E}z_i = 0_{d_x}$ as $\mathsf{E}z_i = 0_{d_z}$. Then explore the first term in $\widehat{b}_c^{(m+1)}$ it follows

$$(M_{\tilde{x}\tilde{x},n}\psi)^{-1}c\mathsf{f}_u(c)n^{-1/2}\sum_{i=1}^n \tilde{x}_{in}\zeta_c^{+\prime}\Sigma^{1/2}\widehat{b}_c^{(m)} \overset{\mathsf{P}}{\to} (M_{\tilde{x}\tilde{x}}\psi)^{-1}c\mathsf{f}_u(c)0_{d_x}\zeta_c^{+\prime}\Sigma^{1/2}\widehat{b}_c^{(m)} = 0_{d_x},$$

uniformly in $c_+ \leq c < \infty$. Uniformity requires $\sup_{c\in\mathbb{R}_+} c|\zeta_c^+|\mathsf{f}_u(c) < \infty$ by Assumption 3.1(ia) and $\widehat{b}_c^{(m)} = \mathrm{O}_{\mathsf{P}}(1)$. Also notice $\zeta_c^+ = 0_{d_x}$ if $\mathsf{E}x_iu_i = 0_{d_x}$ and then $\mathsf{E}r_iu_i = 0_{d_x}$ or under normality assumption (2.5) for $(u_i/\sigma, \Sigma^{-1/2}r_i)$ so the above term also vanishes in these two situations. Then we obtain

$$\widehat{b}_c^{(m+1)} = \frac{2c\mathsf{f}_u(c)}{\psi}\widehat{b}_c^{(m)} + (M_{\tilde{x}\tilde{x},n}\psi)^{-1}\sum_{i=1}^n \tilde{x}_{in}u_i 1_{(|u_i|\leq\sigma c)} + R_\beta(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where $R_\beta(a,b,c)$ vanishes uniformly. A key to above derivation is that c is bounded away from 0 and $M_{\tilde{x}\tilde{x}} > 0$ by Assumption 3.1(iva) so $\psi, M_{\tilde{x}\tilde{x}}\psi > 0$.

Secondly for any $m \in [0,\infty)$ we get an expression for $(\widehat{\sigma}_c^{(m+1)})^2$ in (2.12). By Taylor expansion, first note that

$$n^{1/2}(\widehat{\sigma}_c^{(m+1)} - \sigma) = \frac{1}{2\sigma}n^{1/2}\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\} + n^{-1/2}\mathrm{O}[n\{(\widehat{\sigma}_c^{(m+1)})^2 - \sigma^2\}^2].$$

Then apply arguments as above to get

$$\widehat{a}_c^{(m+1)} = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}\widehat{a}_c^{(m)} + \frac{\sigma}{2\tau_2^c}n^{-1/2}\sum_{i=1}^n(\frac{u_i^2}{\sigma^2} - \varsigma_c^2)1_{(|u_i|\leq\sigma c)} + \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\tau_2^c}\widehat{b}_c^{(m)} + R_\sigma(\widehat{a}_c^{(m)}, \widehat{b}_c^{(m)}, c),$$

where the remainder term $R_\sigma(a,b,c)$ also vanishes uniformly. Notice for analysis of $\widehat{a}_c^{(m+1)}$ we further require the moment condition $\mathsf{E}|x_iu_i|^2 < \infty$ and $\mathsf{E}|x_ix_i'|^2 < \infty$. Since

x_i is correlated with u_i in the IV context and so apply Cauchy-Schwarz inequality to the above product moment, then $\mathsf{E}|x_i u_i|^2 = \mathsf{E}|x_i|^2|u_i|^2 \leq (\mathsf{E}|x_i|^4)^{1/2}(\mathsf{E}|u_i|^4)^{1/2}$. Also note $\mathsf{E}|x_i x_i'|^2 \leq \mathsf{E}|x_i|^4$. Thus we need $\mathsf{E}|u_i|^4 < \infty$, $\mathsf{E}|r_i|^4 < \infty$, $\mathsf{E}|z_i|^4 < \infty$ by Assumption 3.1(ia, iia, ivc). ∎

Remark C.2. *For any $m \in [0, \infty)$ Algorithm 2.1 applies indicators $v_{i,c}^{(m)}$ in (2.9) to choose non-outlying observations for both the first stage regression (2.2) and structural equation (2.1), then updated estimators $\widehat{\Pi}_c^{(m+1)}$, $\widehat{\beta}_c^{(m+1)}$, $(\widehat{\sigma}_c^{(m+1)})^2$ are obtained based on the same sub-sample selected, see (2.10), (2.11), (2.12). The equality (C.4) comes from the fact that the sub-sample chosen for non-outlying observations is the same for (2.2) and (2.1), while it is significant for proving consistency of β estimator. Otherwise without (C.4), we can only get*

$$\widehat{\beta}_c^{(m+1)} = (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i z_i' v_{i,c}^{(m)} \widehat{\Pi}_c^{(m+1)})^{-1} (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i x_i' v_{i,c}^{(m)}) \beta + (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i z_i' v_{i,c}^{(m)} \widehat{\Pi}_c^{(m+1)})^{-1} (\widehat{\Pi}_c^{(m+1)\prime} \sum_{i=1}^n z_i u_i v_{i,c}^{(m)}),$$

and the first product term before β does not reduce to the identity matrix so form the bias term, whereas the second product moment is treated by the same argument in the proof of Theorem 3.2. Thus the estimator for β is inconsistent if two different sub-samples are used for running the first stage and structural equation.

Remark C.3. *The factor ς_c^2 in (2.8) needs to be adjusted in the variance estimator (2.12) to ensure the consistent property. For any $m \in [0, \infty)$ observe the above 1-step stochastic expansion of variance estimator $(\widehat{\sigma}_c^{(m+1)})^2$ multiplied by the rate $n^{-1/2}$*

$$\widehat{\sigma}_c^{(m+1)} - \sigma = \frac{c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c}(\widehat{\sigma}_c^{(m)} - \sigma) + \frac{\sigma}{2\tau_2^c} n^{-1} \sum_{i=1}^n (\frac{u_i^2}{\sigma^2} - \varsigma_c^2) 1_{(|u_i| \leq \sigma c)} + \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\tau_2^c}(\widehat{\beta}_c^{(m)} - \beta) + \mathrm{o}_{\mathsf{P}}(n^{-1/2}).$$

Since assume tightness of m step estimator, consistency holds such that $\widehat{\sigma}_c^{(m)} - \sigma$, $\widehat{\beta}_c^{(m)} - \beta$ are $\mathrm{o}_{\mathsf{P}}(1)$. Furthermore coefficients before them are bounded uniformly in $c \in [c_+, \infty)$ by Assumption 3.1(ia). Then the first and third terms in the above expansion converges to zero in probability. Thus uniform consistency of variance estimator $(\widehat{\sigma}_c^{(m+1)})^2$ requires the second kernel term vanishes. Bias correction factor ς_c^2 in fact appears after minus sign in the bracket in the kernel term, which lead to $n^{-1}\sum_{i=1}^n (u_i^2/\sigma^2 - \varsigma_c^2) 1_{(|u_i| \leq \sigma c)} \overset{\mathsf{P}}{\to} 0$ uniformly in c by Uniform Law of Large Numbers, see the argument in the proof of Lemma C.5.

The next proof is to establish the tightness result.

Proof of Theorem 3.3. Due to Assumption 3.1$(ia, iia, iiia, iv)$, Theorem 3.2 demonstrates for any $m \in [0, \infty)$

$$\widehat{\theta}_c^{(m+1)} = \Gamma_c \widehat{\theta}_c^{(m)} + \Lambda_c + R_\theta(\widehat{\theta}_c^{(m)}, c), \tag{C.6}$$

where the remainder term satisfies $\sup_{c_+ \leq c < \infty} \sup_{|\theta| \leq B} |R_\theta(\theta, c)| = \mathsf{o}_\mathsf{P}(1)$ and

$$\widehat{\theta}_c^{(m)} = \begin{pmatrix} \widehat{b}_c^{(m)} \\ \widehat{a}_c^{(m)} \end{pmatrix} = \begin{Bmatrix} n^{1/2}(\widehat{\beta}_c^{(m)} - \beta) \\ n^{1/2}(\widehat{\sigma}_c^{(m)} - \sigma) \end{Bmatrix}, \quad \Gamma_c = \begin{Bmatrix} \frac{2c\mathsf{f}_u(c)}{\psi} I_{d_x} & 0_{d_x} \\ \frac{(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\tau_2^c} & \frac{c(c^2-\varsigma_c^2)\mathsf{f}_u(c)}{\tau_2^c} \end{Bmatrix}, \tag{C.7}$$

$$\Lambda_c = \begin{Bmatrix} (M_{\tilde{x}\tilde{x},n}\psi)^{-1} & 0_{d_x} \\ 0'_{d_x} & \frac{\sigma}{2\tau_2^c} \end{Bmatrix} \sum_{i=1}^n \begin{Bmatrix} \tilde{x}_{in} u_i \\ n^{-1/2}(\frac{u_i^2}{\sigma^2} - \varsigma_c^2) \end{Bmatrix} 1_{(|u_i| \leq \sigma c)}. \tag{C.8}$$

Apply the autoregressive equation (C.6) recursively to obtain the representation

$$\widehat{\theta}_c^{(m+1)} = \Gamma_c^{m+1} \widehat{\theta}_c^{(0)} + \sum_{l=0}^m \Gamma_c^l \{\Lambda_c + R_\theta(\widehat{\theta}_c^{(m-l)}, c)\}. \tag{C.9}$$

Note the spectral norm of a matrix is compatible with Euclidean norm of a vector so $|Mx| \leq |M||x|$, and use the triangle inequality to get

$$|\widehat{\theta}_c^{(m+1)}| \leq |\Gamma_c^{m+1}||\widehat{\theta}_c^{(0)}| + \{|\Lambda_c| + \max_{0 \leq l \leq m} |R_\theta(\widehat{\theta}_c^{(l)}, c)|\} \sum_{l=0}^m |\Gamma_c^l|.$$

Notice Γ_c has d_x number of eigenvalues $2c\mathsf{f}_u(c)/\psi$ and one eigenvalue $c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)/\tau_2^c$, and (3.1) in Remark 3.2 shows both of them are strictly bounded by one, so we have

$$\sup_{c_+ \leq c < \infty} \max |\text{eigen}(\Gamma_c)| < 1. \tag{C.10}$$

The fact, that the spectral radius of autoregressive coefficient matrix Γ_c is bounded by one as shown in (C.10), is essential to build tightness and fixed point results for the iterative system (C.6). Then because of (C.10) two equalities follow

$$\sum_{l=0}^m \Gamma_c^l = (I_{d_x+1} - \Gamma_c)^{-1}(I_{d_x+1} - \Gamma_c^{m+1}) = (I_{d_x+1} - \Gamma_c^{m+1})(I_{d_x+1} - \Gamma_c)^{-1}. \tag{C.11}$$

When $m \to \infty$, see 5.4(3b) in Lütkepohl (1996), it holds

$$\Gamma_c^{m+1} \to 0_{(d_x+1)\times(d_x+1)}, \qquad \sum_{l=0}^\infty \Gamma_c^l = (I_{d_x+1} - \Gamma_c)^{-1}. \tag{C.12}$$

See Varga (2000, Theorem 3.4), Gelfand's formula gives

$$\lim_{m \to \infty} |\Gamma_c^m|^{1/m} = \max |\text{eigen}(\Gamma_c)|. \tag{C.13}$$

Then (C.13) implies for some ω such that $\sup_{c_+ \leq c < \infty} \max |\text{eigen}(\Gamma_c)| < \omega < 1$ there exists $m_0 > 0$ so for all $m > m_0$ we have

$$\sup_{c_+ \leq c < \infty} |\Gamma_c^m| < \omega^m < 1. \tag{C.14}$$

This with the equality in (C.12) implies for some $1 < B_0 < \infty$

$$\sup_{0\le m<\infty}\sup_{c_+\le c<\infty}|\Gamma_c^m| < B_0, \qquad \sup_{c_+\le c<\infty}|(I_{d_x+1}-\Gamma_c)^{-1}| \le \sum_{l=0}^{\infty}\sup_{c_+\le c<\infty}|\Gamma_c^l| < B_0. \tag{C.15}$$

Use (C.15) to show for all $m \in [0,\infty)$

$$|\widehat{\theta}_c^{(m+1)}| < B_0\{|\widehat{\theta}_c^{(0)}| + |\Lambda_c| + \max_{0\le l\le m}|R_\theta(\widehat{\theta}_c^{(l)}, c)|\}. \tag{C.16}$$

Assumption 3.1(v) with $\eta = 1/4$ guarantees boundedness of $\widehat{\theta}_c^{(0)}$, while the kernel Λ_c is tight by Theorem 4.7 using Assumption 3.1(ia, ivc). Thus, for all $\epsilon, \delta > 0$ there exist $n_0, \Theta_0 > 0$ so that for all $n > n_0$ the set

$$\mathcal{A}_n = \{B_0\sup_{c_+\le c<\infty}(|\widehat{\theta}_c^{(0)}| + |\Lambda_c|) \le \Theta_0/3, B_0\sup_{c_+\le c<\infty}\sup_{|\theta|\le\Theta_0}|R_\theta(\theta, c)| < \delta/2\} \tag{C.17}$$

has probability larger than $1-\epsilon$.

Mathematical induction over m is used to show $\sup_{0\le m<\infty}\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(m)}| \le \Theta_0$ on the set $\mathcal{A}_n$. For $m = 0$ as induction starts, $\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(0)}| \le B_0^{-1}\Theta_0/3 < \Theta_0$ holds since $B_0 > 1$. The induction assumption is that $\sup_{0\le l\le m}\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(l)}| \le \Theta_0$. This implies $B_0\max_{0\le l\le m}|R_\theta(\widehat{\theta}_c^{(l)}, c)| < \delta/2$, and then the bound in (C.16) becomes $\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(m+1)}| < 2\Theta_0/3 + \delta/2 < \Theta_0$ so that $\sup_{0\le l\le m+1}\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(l)}| \le \Theta_0$. ■

Proof of Corollary 3.1. Tightness of initial estimators for structural parameters β, σ^2 in (2.1) implicitly assumes consistency of the initial estimator for location parameter Π in the first stage regression (2.2). Combine Theorem 3.3 and 3.1, then consistency of $\widehat{\Pi}_c^{(m)}$ follows uniformly in $m \in [0,\infty)$ and $c \in [c_+,\infty)$. ■

With tightness results in Theorem 3.3, one-step expansion in Theorem 3.2 can be applied recursively to obtain Theorem 3.4.

Proof of Theorem 3.4. By Assumption 3.1, Theorem 3.2 shows the one-step expansion (C.6). Then apply it recursively to obtain (C.9) so for any $m \in [0,\infty)$

$$\widehat{\theta}_c^{(m+1)} = \Gamma_c^{m+1}\widehat{\theta}_c^{(0)} + \sum_{l=0}^{m}\Gamma_c^l\Lambda_c + \sum_{l=0}^{m}\Gamma_c^l R_\theta(\widehat{\theta}_c^{(m-l)}, c),$$

where $\sup_{c_+\le c<\infty}\sup_{|\theta|\le B}|R_\theta(\theta, c)| = \mathrm{o}_{\mathsf{P}}(1)$. As the spectral radius of Γ_c is bounded by one, see (C.10), then (C.15) shows for $1 < B_0 < \infty$

$$\sup_{0\le m<\infty}\sup_{c_+\le c<\infty}|\Gamma_c^m| < B_0, \qquad \sup_{c_+\le c<\infty}|\sum_{l=0}^{m}\Gamma_c^l| \le \sup_{c_+\le c<\infty}\sum_{l=0}^{m}|\Gamma_c^l| \le \sum_{l=0}^{\infty}\sup_{c_+\le c<\infty}|\Gamma_c^l| < B_0.$$

Further with tightness $\sup_{0\le m<\infty}\sup_{c_+\le c<\infty}|\widehat{\theta}_c^{(m)}| = \mathrm{O}_{\mathsf{P}}(1)$ shown in Theorem 3.3, the third term in the representation of $\widehat{\theta}_c^{(m+1)}$ vanishes, then apply the second equality in (C.11) so uniformly in $c \in [c_+,\infty)$ we have for any $m \in [0,\infty)$

$$\widehat{\theta}_c^{(m+1)} = \Gamma_c^{m+1}\widehat{\theta}_c^{(0)} + (I_{d_x+1} - \Gamma_c^{m+1})(I_{d_x+1} - \Gamma_c)^{-1}\Lambda_c + \mathrm{o}_{\mathsf{P}}(1). \tag{C.18}$$

See Γ_c in (C.7), by matrix manipulations we get

$$\Gamma_c^{m+1} = \begin{pmatrix} \varrho_{\beta\beta,c}^{(m+1)} I_{d_x} & 0_{d_x} \\ \varrho_{\sigma\beta,c}^{(m+1)\prime} & \varrho_{\sigma\sigma,c}^{(m+1)} \end{pmatrix}, (I_{d_x+1} - \Gamma_c^{m+1})(I_{d_x+1} - \Gamma_c)^{-1} = \begin{pmatrix} \psi \varrho_{\beta\tilde{x}u,c}^{(m+1)} I_{d_x} & 0_{d_x} \\ \psi \varrho_{\sigma\tilde{x}u,c}^{(m+1)\prime} & \frac{2\tau_2^c}{\sigma} \varrho_{\sigma uu,c}^{(m+1)} \end{pmatrix},$$

see expressions for $\varrho_{\beta\beta,c}^{(m+1)}$, $\varrho_{\beta\tilde{x}u,c}^{(m+1)}$, $\varrho_{\sigma\sigma,c}^{(m+1)}$, $\varrho_{\sigma uu,c}^{(m+1)}$, $\varrho_{\sigma\beta,c}^{(m+1)}$, $\varrho_{\sigma\tilde{x}u,c}^{(m+1)}$ in Theorem 3.4. Substitute these and $\widehat{\theta}_c^{(0)}$, Λ_c into (C.18), see $\widehat{\theta}_c^{(0)}$ in (C.7) and Λ_c in (C.8), then the expression for $\widehat{\theta}_c^{(m+1)}$ is established. ∎

The next step is to prove the fixed point theorem when iteration step becomes sufficiently large.

Proof of Theorem 3.5. Since the spectral radius of Γ_c is strictly smaller than one, see (C.10), when $m \to \infty$ we can apply (C.12) to the recursive representation (C.18) shown in Theorem 3.4 by Assumption 3.1, then as $n \to \infty$ and uniformly in $c \in [c_+, \infty)$ we have the fixed point

$$\widehat{\theta}_c^* = \widehat{\theta}_c^{(\infty)} = (I_{d_x+1} - \Gamma_c)^{-1}\Lambda_c. \tag{C.19}$$

See Γ_c in (C.7), we get

$$(I_{d_x+1} - \Gamma_c)^{-1} = \begin{bmatrix} \frac{\psi}{\psi - 2c\mathsf{f}_u(c)} I_{d_x} & 0_{d_x} \\ \frac{\psi(c^2 - \varsigma_c^2)\mathsf{f}_u(c)\zeta_c^{-\prime}\Sigma^{1/2}}{2\{\psi - 2c\mathsf{f}_u(c)\}\{\tau_2^c - c(c^2-\varsigma_c^2)\mathsf{f}_u(c)\}} & \frac{\tau_2^c}{\tau_2^c - c(c^2 - \varsigma_c^2)\mathsf{f}_u(c)} \end{bmatrix}.$$

Then substitute this and Λ_c into (C.19), see Λ_c in (C.8), so the expression for the fixed point $\widehat{\theta}_c^*$ is demonstrated.

Substitute (C.9) and (C.19) into the deviation $\widehat{\Delta}_c^{(m+1)} = \widehat{\theta}_c^{(m+1)} - \widehat{\theta}_c^*$, then apply the second equality in (C.12) so uniformly in $c \in [c_+, \infty)$ we have for any $m \in [0, \infty)$

$$\widehat{\Delta}_c^{(m+1)} = \Gamma_c^{m+1}\{\widehat{\theta}_c^{(0)} - (I_{d_x+1} - \Gamma_c)^{-1}\Lambda_c\} + \sum_{l=0}^{m} \Gamma_c^l R_\theta(\widehat{\theta}_c^{(m-l)}, c).$$

To bound $\widehat{\Delta}_c^{(m+1)}$, use the triangle inequality and $|Mx| \le |M||x|$ to get

$$|\widehat{\Delta}_c^{(m+1)}| \le |\Gamma_c^{m+1}|\{|\widehat{\theta}_c^{(0)}| + |(I_{d_x+1} - \Gamma_c)^{-1}||\Lambda_c|\} + \max_{0\le l\le m} |R_\theta(\widehat{\theta}_c^{(l)}, c)| \sum_{l=0}^{m} |\Gamma_c^l|.$$

Now consider the large m such that $m > m_0$ then (C.14) can be used, so along with the second inequality in (C.15) it follows

$$|\widehat{\Delta}_c^{(m+1)}| < \omega^{m+1}(|\widehat{\theta}_c^{(0)}| + B_0|\Lambda_c|) + B_0 \max_{0\le l\le m} |R_\theta(\widehat{\theta}_c^{(l)}, c)|.$$

On the set $\mathcal{A}_n$ as in (C.17), Theorem 3.3 shows $\sup_{0\le m<\infty} \sup_{c_+\le c<\infty} |\widehat{\theta}_c^{(m)}| \le \Theta_0$, so

$$|\widehat{\Delta}_c^{(m+1)}| < \omega^{m+1}(B_0^{-1}\Theta_0/3 + \Theta_0/3) + \delta/2 < \omega^{m+1}\Theta_0 + \delta/2,$$

where the second inequality follows by $1 < B_0 < \infty$. As $0 < \omega < 1$, ω^{m+1} declines exponentially so m_0 can be chosen sufficiently large that for $m > m_0$ then $\omega^{m+1}\Theta_0 < \delta/2$. Thus for $n > n_0$, $m > m_0$ we have $\mathsf{P}(\sup_{c_+\le c<\infty} |\widehat{\Delta}_c^{(m+1)}| < \delta) > 1 - \epsilon$. ∎

C.3 Results for Algorithm 2.2 and 2.3

We first present a lemma to describe the limiting distribution of kernels.

Lemma C.6. *Suppose Assumption 3.1(ia, iva, ivc) holds. Then as $n \to \infty$ we have for any $c \in [c_+, \infty)$*

$$\sum_{i=1}^{n} \begin{pmatrix} \tilde{x}_{in} u_i \\ \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)} \end{pmatrix} \overset{\mathsf{D}}{\to} \mathsf{N} \left\{ \begin{pmatrix} 0_{d_x} \\ 0_{d_x} \end{pmatrix}, \sigma^2 \tau_2^c \begin{pmatrix} \frac{1}{\tau_2^c} M_{\tilde{x}\tilde{x}} & M_{\tilde{x}\tilde{x}} \\ M_{\tilde{x}\tilde{x}} & M_{\tilde{x}\tilde{x}} \end{pmatrix} \right\}.$$

Proof of Theorem C.6. As $\tilde{x}_i = \Pi' z_i$ and z_i is independent of u_i by Assumption 3.1. Then $\mathsf{E}\tilde{x}_i u_i = 0_{d_x}$ and $\mathsf{Var}(\tilde{x}_i u_i) = \mathsf{E}\tilde{x}_i \tilde{x}_i' u_i^2 = \mathsf{E}u_i^2 \mathsf{E}\tilde{x}_i \tilde{x}_i' = \sigma^2 M_{\tilde{x}\tilde{x}}$. Furthermore we have $\mathsf{E}\tilde{x}_i u_i 1_{(|u_i| \leq \sigma c)} = \mathsf{E}\tilde{x}_i \mathsf{E}u_i 1_{(|u_i| \leq \sigma c)} = 0_{d_x}$ and

$$\mathsf{Var}(\tilde{x}_i u_i 1_{(|u_i| \leq \sigma c)}) = \mathsf{E}\tilde{x}_i \tilde{x}_i' u_i^2 1_{(|u_i| \leq \sigma c)} = \mathsf{E}u_i^2 1_{(|u_i| \leq \sigma c)} \mathsf{E}\tilde{x}_i \tilde{x}_i' = \sigma^2 \tau_2^c M_{\tilde{x}\tilde{x}}.$$

Finally we calculate $\mathsf{Cov}(\tilde{x}_i u_i, \tilde{x}_i u_i 1_{(|u_i| \leq \sigma c)}) = \mathsf{E}\tilde{x}_i \tilde{x}_i' u_i^2 1_{(|u_i| \leq \sigma c)} = \sigma^2 \tau_2^c M_{\tilde{x}\tilde{x}}$. Note $\tilde{x}_{in} = n^{-1/2} \tilde{x}_i$, so apply CLT to obtain for any $c \in [c_+, \infty)$

$$n^{-1/2} \sum_{i=1}^{n} \begin{pmatrix} \tilde{x}_i u_i \\ \tilde{x}_i u_i 1_{(|u_i| \leq \sigma c)} \end{pmatrix} \overset{\mathsf{D}}{\to} \mathsf{N} \left\{ \begin{pmatrix} 0_{d_x} \\ 0_{d_x} \end{pmatrix}, \begin{pmatrix} \sigma^2 M_{\tilde{x}\tilde{x}} & \sigma^2 \tau_2^c M_{\tilde{x}\tilde{x}} \\ \sigma^2 \tau_2^c M_{\tilde{x}\tilde{x}} & \sigma^2 \tau_2^c M_{\tilde{x}\tilde{x}} \end{pmatrix} \right\}. \qquad \blacksquare$$

The next step is to provide the expansion and limiting distribution of the full sample two stage least squares estimator.

Proof of Lemma 3.1. Substitute (2.2) into the location estimator for Π to obtain

$$\widetilde{\Pi} = \Pi + M_{zz,n}^{-1} n^{-1/2} \sum_{i=1}^{n} z_{in} r_i' \overset{\mathsf{P}}{\to} \Pi + M_{zz}^{-1} \mathsf{E} z_i r_i' = \Pi + M_{zz}^{-1} 0_{d_z \times d_x} = \Pi,$$

where the probability limit follows by Assumption 3.1(iva) that $M_{zz,n} \overset{\mathsf{P}}{\to} M_{zz} > 0$, LLN, Slutsky's theorem, and $\mathsf{E} z_i r_i' = 0_{d_z \times d_x}$. Notice LLN requires moment conditions $\mathsf{E}|r_i|, \mathsf{E}|z_i| < \infty$ by Assumption 3.1(iia, ivc). We have $\widetilde{\Pi}' \sum_{i=1}^{n} z_i z_i' \widetilde{\Pi} = \widetilde{\Pi}' \sum_{i=1}^{n} z_i x_i'$, see (C.4), then apply this and structural equation (2.1) to get

$$n^{1/2}(\widetilde{\beta} - \beta) = (\widetilde{\Pi}' M_{zz,n} \widetilde{\Pi})^{-1} (\widetilde{\Pi}' \sum_{i=1}^{n} z_{in} u_i) = M_{\tilde{x}\tilde{x},n}^{-1} \sum_{i=1}^{n} \tilde{x}_{in} u_i + \mathsf{o}_{\mathsf{P}}(1),$$

where the second equality is due to consistency of $\widetilde{\Pi}$. Thus the limiting distribution follows by Lemma C.6 and Assumption 3.1(iva) that $M_{\tilde{x}\tilde{x},n} \overset{\mathsf{P}}{\to} M_{\tilde{x}\tilde{x}} > 0$. $\blacksquare$

Then weak convergence results are proved for Algorithm 2.2, 2.3.

Proof of Theorem 3.6. Algorithm 2.2 chooses the full sample two stage least squares as the initial estimator $\widehat{\beta}_c^{(0)}$ in Algorithm 2.1. Thus substitute the asymptotic expansion of $n^{1/2}(\widehat{\beta}_c^{(0)} - \beta) = n^{1/2}(\widetilde{\beta} - \beta)$ in Lemma 3.1 into the recursive equation in Theorem 3.4, then uniformly in $c \in [c_+, \infty)$ we have for any $m \in [0, \infty)$

$$\mathbb{G}_n^{(m+1)}(c) = \begin{pmatrix} \varrho_{\beta\beta,c}^{(m+1)} M_{\tilde{x}\tilde{x},n}^{-1} \\ \varrho_{\beta\tilde{x}u,c}^{(m+1)} M_{\tilde{x}\tilde{x},n}^{-1} \end{pmatrix}' \sum_{i=1}^{n} \begin{pmatrix} \tilde{x}_{in} u_i \\ \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)} \end{pmatrix} + \mathsf{o}_{\mathsf{P}}(1),$$

where expressions of $\varrho^{(m+1)}_{\beta\beta,c}$, $\varrho^{(m+1)}_{\beta\tilde{x}u,c}$ are shown in Theorem 3.4. Then $M_{\tilde{x}\tilde{x},n} \overset{\mathsf{P}}{\to} M_{\tilde{x}\tilde{x}} > 0$, Lemma C.6, and Slutsky's theorem show for any $c \in [c_+, \infty)$

$$\mathbb{G}^{(m+1)}_n(c) \overset{\mathsf{D}}{\to} \begin{pmatrix} \varrho^{(m+1)}_{\beta\beta,c} M^{-1}_{\tilde{x}\tilde{x}} \\ \varrho^{(m+1)}_{\beta\tilde{x}u,c} M^{-1}_{\tilde{x}\tilde{x}} \end{pmatrix}' \mathsf{N}\left\{ \begin{pmatrix} 0_{d_x} \\ 0_{d_x} \end{pmatrix}, \sigma^2 \tau^c_2 \begin{pmatrix} \frac{1}{\tau^c_2} M_{\tilde{x}\tilde{x}} & M_{\tilde{x}\tilde{x}} \\ M_{\tilde{x}\tilde{x}} & M_{\tilde{x}\tilde{x}} \end{pmatrix} \right\}.$$

Since a transformation of multivariate normal is still normal, it follows

$$\mathbb{G}^{(m+1)}_n(c) \overset{\mathsf{D}}{\to} \mathsf{N}[0_{d_x}, \{(\varrho^{(m+1)}_{\beta\beta,c})^2 + 2\tau^c_2 \varrho^{(m+1)}_{\beta\beta,c} \varrho^{(m+1)}_{\beta\tilde{x}u,c} + \tau^c_2 (\varrho^{(m+1)}_{\beta\tilde{x}u,c})^2\} \sigma^2 M^{-1}_{\tilde{x}\tilde{x}}].$$

Theorem 3.3 demonstrates the process $\mathbb{G}^{(m+1)}_n$ is tight for any $m \in [0, \infty)$ thus

$$\mathbb{G}^{(m+1)}_n \rightsquigarrow \mathbb{G}^{(m+1)},$$

where the weak limit $\mathbb{G}^{(m+1)}$ is a zero mean Gaussian process with the variance

$$\mathsf{Var}\{\mathbb{G}^{(m+1)}(c)\} = \{(\varrho^{(m+1)}_{\beta\beta,c})^2 + 2\tau^c_2 \varrho^{(m+1)}_{\beta\beta,c} \varrho^{(m+1)}_{\beta\tilde{x}u,c} + \tau^c_2 (\varrho^{(m+1)}_{\beta\tilde{x}u,c})^2\} \sigma^2 M^{-1}_{\tilde{x}\tilde{x}}. \qquad \blacksquare$$

Proof of Corollary 3.2. This is a special case of Theorem 3.6 when $m = 0$ such that $\varrho^{(1)}_{\beta\beta,c} = 2c\mathsf{f}_u(c)/\psi$, $\varrho^{(1)}_{\beta\tilde{x}u,c} = \psi^{-1}$, see (3.2). ■

Proof of Theorem 3.7. The first step updated estimator for Π is expressed as

$$\widehat{\Pi}^{(1)}_c - \Pi = (\sum_{j=1,2} \frac{n_{3-j}}{n} \sum_{i \in \mathcal{I}_{3-j}} z_{in_{3-j}} z'_{in_{3-j}} 1_{(|y_i - x'_i \widehat{\beta}_j| \leq \widehat{\sigma}_j c)})^{-1}$$
$$(\sum_{j=1,2} \frac{n_{3-j}}{n} n^{-1/2}_{3-j} \sum_{i \in \mathcal{I}_{3-j}} z_{in_{3-j}} r'_i 1_{(|y_i - x'_i \widehat{\beta}_j| \leq \widehat{\sigma}_j c)}).$$

Assumption 3.1(ia, iia, iva, ivc) holds for each of sub-sample $\mathcal{I}_j$, so two stage least squares $\widehat{\beta}_j, \widehat{\sigma}^2_j$ are tight for $j = 1, 2$, see Lemma 3.1. Argue as Theorem 3.1 to obtain

$$\widehat{\Pi}^{(1)}_c - \Pi = (\sum_{i=1}^n z_{in} z'_{in} 1_{(|u_i| \leq \sigma c)})^{-1} (n^{-1/2} \sum_{i=1}^n z_{in} r'_i 1_{(|u_i| \leq \sigma c)}) + \mathsf{o}_\mathsf{P}(1),$$

uniformly in $c \in [c_+, \infty)$. Therefore uniform consistency of $\widehat{\Pi}^{(1)}_c$ immediately follows, see proof of Theorem 3.1.

The updated estimator for β is expressed as

$$n^{1/2}(\widehat{\beta}^{(1)}_c - \beta) = (\widehat{\Pi}^{(1)\prime}_c \sum_{j=1,2} \frac{n_{3-j}}{n} \sum_{i \in \mathcal{I}_{3-j}} z_{in_{3-j}} z'_{in_{3-j}} 1_{(|y_i - x'_i \widehat{\beta}_j| \leq \widehat{\sigma}_j c)} \widehat{\Pi}^{(1)}_c)^{-1}$$
$$(\widehat{\Pi}^{(1)\prime}_c \sum_{j=1,2} \sqrt{\frac{n_{3-j}}{n}} \sum_{i \in \mathcal{I}_{3-j}} z_{in_{3-j}} u_i 1_{(|y_i - x'_i \widehat{\beta}_j| \leq \widehat{\sigma}_j c)}).$$

Argue along the lines of Theorem 3.2, then it follows

$$n^{1/2}(\widehat{\beta}^{(1)}_c - \beta) = (M_{\tilde{x}\tilde{x},n} \psi)^{-1} \{2c\mathsf{f}_u(c) \sum_{j=1,2} \sqrt{\frac{n_{3-j}}{n}} \sqrt{\frac{n_{3-j}}{n_j}} M^{\mathcal{I}_{3-j}}_{\tilde{x}\tilde{x},n_{3-j}} n^{1/2}_j (\widehat{\beta}_j - \beta)$$
$$+ \sum_{i=1}^n \tilde{x}_{in} u_i 1_{(|u_i| \leq \sigma c)}\} + \mathsf{o}_\mathsf{P}(1),$$

uniformly in $c \in [c_+, \infty)$ and where we denote $M^{\mathcal{I}_j}_{\tilde{x}\tilde{x},n_j} = \sum_{i\in\mathcal{I}_j} \tilde{x}_{in_j}\tilde{x}'_{in_j}$ for $j = 1, 2$. Again Assumption 3.1(ia, iia, iva, ivc) holds for each sub-sample $\mathcal{I}_j$, so Lemma 3.1 shows $n_j^{1/2}(\widehat{\beta}_j - \beta) = (M^{\mathcal{I}_j}_{\tilde{x}\tilde{x},n_j})^{-1} \sum_{i\in\mathcal{I}_j} \tilde{x}_{in_j}u_i + \mathsf{o}_\mathsf{P}(1)$ and notice $M^{\mathcal{I}_j}_{\tilde{x}\tilde{x},n_j} \overset{\mathsf{P}}{\to} M_{\tilde{x}\tilde{x}}$ as $n \to \infty$ such that $n_j \to \infty$ for each $j = 1, 2$. Thus we have

$$n^{1/2}(\widehat{\beta}^{(1)}_c - \beta) = \frac{2c\mathsf{f}_u(c)}{\psi} M^{-1}_{\tilde{x}\tilde{x},n} \sum_{j=1,2} \frac{n_{3-j}}{n_j} \sum_{i\in\mathcal{I}_j} \tilde{x}_{in}u_i + (M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in}u_i 1_{(|u_i|\leq\sigma c)} + \mathsf{o}_\mathsf{P}(1),$$

and if assume $n_1 = n_2 = n/2$ it further holds

$$n^{1/2}(\widehat{\beta}^{(1)}_c - \beta) = \frac{2c\mathsf{f}_u(c)}{\psi} M^{-1}_{\tilde{x}\tilde{x},n} \sum_{i=1}^{n} \tilde{x}_{in}u_i + (M_{\tilde{x}\tilde{x},n}\psi)^{-1} \sum_{i=1}^{n} \tilde{x}_{in}u_i 1_{(|u_i|\leq\sigma c)} + \mathsf{o}_\mathsf{P}(1).$$

We find the expansion of $n^{1/2}(\widehat{\beta}^{(1)}_c - \beta)$ for Algorithm 2.3 is the same as Algorithm 2.2, see Corollary 3.2, therefore two algorithms have the identical asymptotics even they apply different initial estimates when running Algorithm 2.1. ■

Finally we establish weak convergence for the fixed point of Algorithm 2.1, 2.2, 2.3.

Proof of Theorem 3.8. Consider Algorithm 2.1, 2.2, 2.3. Let $m \to \infty$ then Theorem 3.5 shows the fixed point for $c \in [c_+, \infty)$

$$\mathbb{G}^*_n(c) = \frac{1}{\psi - 2c\mathsf{f}_u(c)} M^{-1}_{\tilde{x}\tilde{x},n} \sum_{i=1}^{n} \tilde{x}_{in}u_i 1_{(|u_i|\leq\sigma c)}.$$

As $n \to \infty$ then $M_{\tilde{x}\tilde{x},n} \overset{\mathsf{P}}{\to} M_{\tilde{x}\tilde{x}} > 0$ and Lemma C.6 demonstrate finite dimensional convergence

$$\mathbb{G}^*_n(c) \overset{\mathsf{D}}{\to} \mathsf{N}[0_{d_x}, \frac{\tau^c_2}{\{\psi - 2c\mathsf{f}_u(c)\}^2}\sigma^2 M^{-1}_{\tilde{x}\tilde{x}}].$$

Tightness shown in Theorem 3.3 further establishes weak convergence

$$\mathbb{G}^*_n \rightsquigarrow \mathbb{G}^*,$$

where $\mathbb{G}^*$ is a zero mean Gaussian process with variance

$$\mathsf{Var}\{\mathbb{G}^*(c)\} = \frac{\tau^c_2}{\{\psi - 2c\mathsf{f}_u(c)\}^2}\sigma^2 M^{-1}_{\tilde{x}\tilde{x}}.$$ ■

References

[1] Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70, 91-117.

[2] Acemoglu, D., Johnson, S., and Robinson, J.A. (2001). The colonial origins of comparative development: an empirical investigation. *American Economic Review* 91, 1369-1401.

[3] Acemoglu, D., Johnson, S., and Robinson, J.A. (2012). The colonial origins of comparative development: an empirical investigation: reply. *American Economic Review* 102, 3077-3110.

[4] Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J.A. (2017). Democracy does cause growth. *Journal of Political Economy*, forthcoming.

[5] Adichie, J.N. (1967). Estimates of regression parameters based on rank tests. *Annals of Mathematical Statistics* 38, 894-904.

[6] Adler, R.J. and Taylor, J.E. (2009). *Random Fields and Geometry*. New York: Springer.

[7] Albouy, D.Y. (2012). The colonial origins of comparative development: an empirical investigation: comment. *American Economic Review* 102, 3059-3076.

[8] Anderson, T.W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46-63.

[9] Auerbach, A.J., Hassett, K.A., and Oliner, S.D. (1994). Reassessing the social returns to equipment investment. *Quarterly Journal of Economics* 109, 789-802.

[10] Bercu, B. and Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18, 1848-1869.

[11] Berenguer-Rico, V., Johansen, S., and Nielsen, B. (2018). Some new results on marked and weighted empirical processes of residuals. *Working Paper*.

[12] Berenguer-Rico, V. and Nielsen, B. (2018). Marked and weighted empirical processes of residuals with applications to robust regressions. *Working Paper*.

[13] Berenguer-Rico, V. and Wilms, I. (2018). White heteroscedasticty testing after outlier removal. *Working Paper*.

[14] Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70, 428-434.

[15] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

[16] Cavaliere, G. and Georgiev, I. (2013). Exploiting infinite variance through dummy variables in nonstationary autoregression. *Econometric Theory* 29, 1162-1195.

[17] Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects. *Econometrica* 73, 245-261.

[18] Chhaochharia, V. and Grinstein, Y. (2009). CEO compensation and board structure. *Journal of Finance* 64, 231-261.

[19] Cragg, J.G. and Donald, S.G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9, 222-240.

[20] De Long, J.B. and Summers, L.H. (1991). Equipment investment and economic growth. *Quarterly Journal of Economics* 106, 445-502.

[21] De Long, J.B. and Summers, L.H. (1994). Equipment investment and economic growth: reply. *Quarterly Journal of Economics* 109, 803-807.

[22] Dollinger, M.B. and Staudte, R.G. (1991). Influence functions of iteratively reweighted least squares estimators. *Journal of the American Statistical Association* 86, 709-716.

[23] Donsker, M.D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics* 23, 277-281.

[24] Doob, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Annals of Mathematical Statistics* 20, 393-403.

[25] Doornik, J.A. (2009). Autometrics. In Castle, J.L. and Shephard, N. (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honor of David F. Hendry*, pp. 88-121. Oxford: Oxford University Press.

[26] Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometrics Journal* 12, 367-381.

[27] Fabrizio, K.R., Rose, N.L., and Wolfram, C.D. (2007). Do markets reduce costs? Assessing the impact of regulatory restructuring on US electric generation efficiency. *American Economic Review* 97, 1250-1277.

[28] Freedman, D.A. (1975). On tail probabilities for martingales. *Annals of Probability* 3, 100-118.

[29] Guthrie, K., Sokolowsky, J., and Wan, K.M. (2012). CEO compensation and board structure revisited. *Journal of Finance* 67, 1149-1168.

[30] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics*. New York: Wiley.

[31] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

[32] Hansen, L.P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics* 14, 262-280.

[33] Hausman, J.A. (1978). Specification tests in econometrics. *Econometrica* 46, 1251-1271.

[34] Hendry, D.F. (1999). An econometric analysis of US food expenditure, 1931-1989. In Magnus, J.R. and Morgan, M.S. (eds.) *Methodology and Tacit Knowledge: Two Experiments in Econometrics*, pp. 341-361. New York: Wiley.

[35] Hendry, D.F. and Doornik, J.A. (2014). *Empirical Model Discovery and Theory Evaluation*. Cambridge MA: MIT Press.

[36] Hendry, D.F., Johansen, S., and Santos, C. (2008). Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 23, 317-335. Erratum 337-339.

[37] Hendry, D.F. and Santos, C. (2010). An automatic test of super exogeneity. In Bollerslev, T., Russell, J.R., and Watson, M.W. (eds.) *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, pp. 164-193. Oxford: Oxford University Press.

[38] Herndon, T., Ash, M., and Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics* 38, 257-279.

[39] Honoré, B.E. and Hu, L. (2004). On the performance of some robust instrumental variables estimators. *Journal of Business and Economic Statistics* 22, 30-39.

[40] Honoré, B.E. and Powell, J.L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 64, 241-278.

[41] Hoover, K.D. and Perez, S.J. (1999). Data mining reconsidered: encompassing and the general-to-specific approach to specification search. *Econometrics Journal* 2, 167-191.

[42] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.

[43] Imbens, G.W. and Angrist, J.D. (1994). Identification and estimation of local average treatment effects. *Econometrica* 62, 467-475.

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

[45] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* 40, 633-643.

[46] Jiao, X. (2018). Uniform convergence of marked empirical processes in location and scale model. *Working Paper*.

[47] Jiao, X. and Nielsen, B. (2017). Asymptotic analysis of iterated 1-step Huber-skip M-estimators with varying cut-offs. In Antoch, J., Jurečková, J., Maciak, M., and Pešta, M. (eds.) *Analytical Methods in Statistics: AMISTAT, Prague, November 2015*. Springer Proceedings in Mathematics and Statistics 193, 23-52.

[48] Jiao, X. and Pretis, F. (2018). Testing the presence of outliers to assess misspecfication in regression models. *Working Paper*.

[49] Johansen, S. and Nielsen, B. (2009). An analysis of the indicator saturation estimator as a robust regression estimator. In Castle, J.L. and Shephard, N. (eds.) *The Methodology and Practice of Econometrics: A Festschrift in Honor of David F. Hendry*, pp. 1-36. Oxford: Oxford University Press.

[50] Johansen, S. and Nielsen, B. (2013). Asymptotic theory for iterated one-step Huber-skip estimators. *Econometrics* 1, 53-70.

[51] Johansen, S. and Nielsen, B. (2016a). Analysis of the Forward Search using some new results for martingales and empirical processes. *Bernoulli* 22, 1131-1183.

[52] Johansen, S. and Nielsen, B. (2016b). Asymptotic theory of outlier detection algorithms for linear time series regression models (with discussion). *Scandinavian Journal of Statistics* 43, 321-381.

[53] Johansen, S. and Nielsen, B. (2018). Boundedness of M-estimators for multiple linear regression in time series. *Econometric Theory*, forthcoming.

[54] Jurečková, J., Sen, P.K., and Picek, J. (2013). *Methodology in Robust and Nonparametric Statistics*. Boca Raton, FL: CRC Press.

[55] Kaji, T. (2018). Switching to the new norm: from heuristics to formal tests using integrable empirical processes. *Working Paper*.

[56] Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781-1803.

[57] Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica* 46. 33-50.

[58] Kolmogorov, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale Dell'Instituto Italiano Degli Attuari* 4, 83-91.

[59] Koul, H.L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models*, 2nd edition. New York: Springer.

[60] Koul, H.L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Annals of Statistics* 34, 994-1012.

[61] Koul, H.L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Annals of Statistics* 22, 540-562.

[62] Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley.

[63] Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*. Chicester: Wiley.

[64] Maronna, R.A. and Yohai, V.J. (1997). Robust estimation in simultaneous equations models. *Journal of Statistical Planning and Inference* 65, 233-244.

[65] Moreira, M.J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027-1048.

[66] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley.

[67] Nielsen, B. and Qian, M. (2018). Asymptotic properties of the gauge of step-indicator saturation. *Working Paper*.

[68] Reinhart, C.M. and Rogoff, K.S. (2010). Growth in a time of debt. *American Economic Review* 100, 573-578.

[69] Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

[70] Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828-838.

[71] Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393-415.

[72] Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Mathèmatique de l'Université de Moscou* 2(2).

[73] Staiger, D. and Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

[74] Stock, J.H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews, D.W.K. and Stock, J.H. (eds.) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80-108. Cambridge: Cambridge University Press.

[75] Toda, A.A. and Walsh, K.J. (2015). The double power law in consumption and implications for testing euler equations. *Journal of Political Economy* 123, 1177-1200.

[76] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

[77] Varga, R.S. (2000). *Matrix Iterative Analysis*. Berlin: Springer.

[78] Welsh, A.H. and Ronchetti, E. (2002). A journey in single steps: robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference* 103, 287-310.

[79] Wüthrich, K. (2018). A comparison of two quantile models with endogeneity. *Journal of Business and Economic Statistics*, forthcoming.

[80] Young, A. (2017). Channelling Fisher: randomization tests and the statistical insignificance of seemingly significant experimental results. *Working Paper*.

[81] Young, A. (2018). Consistency without inference: instrumental variables in practical application. *Working Paper*.