Université de Montréal

Essays on bootstrap methods in econometrics

par
Maximilien Kaffo Melou

Département de sciences économiques
Faculté des arts et sciences

Thèse présentée à la Faculté des arts et sciences
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en sciences économiques

May, 2014

Université de Montréal
Faculté des arts et sciences

Cette thèse intitulée :

Essays on bootstrap methods in econometrics

présentée par :

Maximilien Kaffo Melou

a été évaluée par un jury composé des personnes suivantes :

Marine Carrasco	président-rapporteur
Sílvia Gonçalves,	directrice de recherche
Benoit Perron,	membre du jury
James Gordon Mackinnon,	examinateur externe
Christian Léger,	représentant du doyen de la FAS

à mes parents, Madeleine Tchinda et Martin Melou

Remerciements

Je serai toujours redevable à ma directrice de thèse, Sílvia Gonçalves. Ses précieux conseils vont me guider pour le reste de ma vie professionnelle. Je n'aurais certainement pas pu terminer cette thèse sans ses encouragements et son constant soutien.

Mes remerciements vont aussi à Benoit Perron et Marine Carrasco, membres de mon comité de thèse, pour leurs précieuses suggestions et critiques constructives. Merci à Russel Davidson et Marc Henry pour l'appui qu'ils m'ont apporté tout au long de ma thèse.

Je suis également très reconnaissant à Sílvia Gonçalves et Wenjie Wang pour m'avoir permis d'être leur co-auteur. Travailler avec eux a été une expérience extraordinaire, et j'ai hâte de continuer à apprendre d'eux dans les années à avenir.

J'ai eu l'honneur de partager ces années avec mes camarades de classe et amis, Anabelle, Cathérine, Guy, Hector, Juste, Louis-Philippe, Vincent de Paul, Walter. J'ai beaucoup appris d'eux. Mes pensées vont aussi à ma famille. En particulier mes parents, Madeleine et Martin, pour leur inlassable soutien et surtout d'avoir toujours cru en moi. Plus spécialement, je tiens à remercier ma merveilleuse épouse Line pour avoir été ce support sur lequel je pouvais m'appuyer quand tout semblait s'obscurcir devant moi.

Enfin, je remercie le Département de sciences économiques de l'Université de Montréal, le Centre Interuniversitaire de Recherche en économie Quantitative (CIREQ), la Faculté des études supérieures et postdoctorales et le

Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC) pour le soutien financier. Mes remerciements vont également à tout le corps professoral et au personnel administratif.

Résumé

Ma thèse est composée de trois essais sur l'inférence par le bootstrap à la fois dans les modèles de données de panel et les modèles à grands nombres de variables instrumentales (VI) dont un grand nombre peut être faible. La théorie asymptotique n'étant pas toujours une bonne approximation de la distribution d'échantillonnage des estimateurs et statistiques de tests, je considère le bootstrap comme une alternative. Ces essais tentent d'étudier la validité asymptotique des procédures bootstrap existantes et quand invalides, proposent de nouvelles méthodes bootstrap valides.

Le premier chapitre (co-écrit avec Sílvia Gonçalves) étudie la validité du bootstrap pour l'inférence dans un modèle de panel de données linéaire, dynamique et stationnaire à effets fixes. Nous considérons trois méthodes bootstrap: le recursive-design bootstrap, le fixed-design bootstrap et le pairs bootstrap. Ces méthodes sont des généralisations naturelles au contexte des panels des méthodes bootstrap considérées par Gonçalves et Kilian (2004) dans les modèles autorégressifs en séries temporelles. Nous montrons que l'estimateur MCO obtenu par le recursive-design bootstrap contient un terme intégré qui imite le biais de l'estimateur original. Ceci est en contraste avec le fixed-design bootstrap et le pairs bootstrap dont les distributions sont incorrectement centrées à zéro. Cependant, le recursive-design bootstrap et le pairs bootstrap sont asymptotiquement valides quand ils sont appliqués à l'estimateur corrigé du biais, contrairement au fixed-design bootstrap. Dans les simulations, le recursive-design bootstrap est la méthode qui produit les meilleurs résultats.

Le deuxième chapitre étend les résultats du pairs bootstrap aux modèles de panel non linéaires dynamiques avec des effets fixes. Ces modèles sont souvent estimés par l'estimateur du maximum de vraisemblance (EMV) qui souffre également d'un biais. Récemment, Dhaene et Johmans (2014) ont proposé la méthode d'estimation split-jackknife. Bien que ces estimateurs ont des approximations asymptotiques normales centrées sur le vrai paramètre, de sérieuses distorsions demeurent à échantillons finis. Dhaene et Johmans (2014) ont proposé le pairs bootstrap comme alternative dans ce contexte sans aucune justification théorique. Pour combler cette lacune, je montre que cette méthode est asymptotiquement valide lorsqu'elle est utilisée pour estimer la distribution de l'estimateur split-jackknife bien qu'incapable d'estimer la distribution de l'EMV. Des simulations Monte Carlo montrent que les intervalles de confiance bootstrap basés sur l'estimateur split-jackknife aident grandement à réduire les distorsions liées à l'approximation normale en échantillons finis. En outre, j'applique cette méthode bootstrap à un modèle de participation des femmes au marché du travail pour construire des intervalles de confiance valides.

Dans le dernier chapitre (co-écrit avec Wenjie Wang), nous étudions la validité asymptotique des procédures bootstrap pour les modèles à grands nombres de variables instrumentales (VI) dont un grand nombre peu être faible. Nous montrons analytiquement qu'un bootstrap standard basé sur les résidus et le bootstrap restreint et efficace (RE) de Davidson et MacKinnon (2008, 2010, 2014) ne peuvent pas estimer la distribution limite de l'estimateur du maximum de vraisemblance à information limitée (EMVIL). La raison principale est qu'ils ne parviennent pas à bien imiter le paramètre qui caractérise l'intensité de l'identification dans l'échantillon. Par conséquent, nous proposons une méthode bootstrap modifiée qui estime de facon convergente cette distribution limite. Nos simulations montrent que la méthode bootstrap modifiée réduit considérablement les distorsions des tests asymptotiques de type Wald (t) dans les échantillons finis, en particulier lorsque le degré d'endogénéité est élevé.

Keywords : bootstrap, données de panel, effets fixes, split-jackknife, instruments faibles, EMVIL, bootstrap restreint et efficace.

Abstract

My dissertation consists of three essays on bootstrap inference in both large panel data models and instrumental variable (IV) models with many instruments and possibly, many weak instruments. Since the asymptotic theory is often not a good approximation to the sampling distribution of test statistics and estimators, I consider the bootstrap as an alternative. These essays try to study the asymptotic validity of existing bootstrap procedures and when they are invalid, to propose new valid bootstrap methods.

The first chapter (co-authored with Sílvia Gonçalves) studies the validity of the bootstrap for inference on a stationary linear dynamic panel data model with individual fixed effects. We consider three bootstrap methods: the recursive-design wild bootstrap, the fixed-design wild bootstrap and the pairs bootstrap. These methods are natural generalizations to the panel context of the bootstrap methods considered by Gonçalves and Kilian (2004) in pure time series autoregressive models. We show that the recursive-design wild bootstrap fixed effects OLS estimator contains a built-in bias correction term that mimics the incidental parameter bias. This is in contrast with the fixed-design wild bootstrap and the pairs bootstrap whose distributions are incorrectly centered at zero. As it turns out, both the recursive-design and the pairs bootstrap are asymptotically valid when applied to the bias-corrected estimator, but the fixed-design bootstrap is not. In the simulations, the recursive-design bootstrap is the method that does best overall.

The second chapter extends our pairwise bootstrap results to dynamic nonlinear panel data models with fixed effects. These models are often estimated with the Maximum Likelihood Estimator (MLE) which also suffers from an incidental parameter bias. Recently, Dhaene and Jochmans (2014) have proposed the split-jackknife estimation method. Although these estimators have asymptotic normal approximations that are centered at the true parameter, important size distortions remain in finite samples. Dhaene and Jochmans (2014) have proposed the pairs bootstrap as an alternative in this context without a theoretical justification. To fill this gap, I show that this method is asymptotically valid when used to estimate the distribution of the half-panel jackknife estimator although it does not consistently estimate the distribution of the MLE. A Monte Carlo experiment shows that bootstrap-based confidence intervals that rely on the half-panel jackknife estimator greatly help to reduce the distortions associated to the normal approximation in finite samples. In addition, I apply this bootstrap method to a canonical model of female-labor participation to construct valid confidence intervals.

In the last chapter (co-authored with Wenjie Wang), we study the asymptotic validity of bootstrap procedures for instrumental variable (IV) models with many weak instruments. We show analytically that a standard residual-based bootstrap and the restricted efficient (RE) bootstrap of Davidson and MacKinnon (2008, 2010, 2014) cannot consistently estimate the limiting distribution of the LIML estimator. The foremost reason is that they fail to adequately mimic the identification strength in the sample. Therefore, we propose a modified bootstrap procedure which consistently estimates this limiting distribution. Our simulations show that the modified bootstrap procedure greatly reduces the distortions associated to asymptotic Wald (t) tests in finite samples, especially when the degree of endogeneity is high.

Keywords : bootstrap, dynamic panel data, fixed effects, incidental parameter bias, half-panel jackknife, weak instruments, LIML, RE bootstrap.

Introduction

This dissertation is a collection of three essays in theoretical and applied econometrics, organized in the form of three chapters. In the three chapters, my focus is on the bootstrap as a method of inference. The first two chapters consider its application to panel data models with individual fixed effects while the last chapter considers Instrumental Variable (IV) models with many instruments and, possibly, many weak instruments.

The asymptotic theory provides an approximation to the sampling distribution of test statistics and estimators. However, it is now well known that for the sample sizes encountered in practice, the asymptotic theory is often not a good approximation. The bootstrap is an alternative method of inference that can be used to approximate the distribution of an estimator or characteristics of that distribution such as a variance or a quantile. It generally provides a better approximation in finite samples than the standard asymptotic theory approximations and is extensively used in applied research, although sometimes without any theoretical foundation. This thesis aims to fill this gap.

In the first chapter, we propose and theoretically justify the application of bootstrap methods for inference in autoregressive panel data models with fixed effects. Whereas the focus of the existing literature has been on bias correcting the standard fixed effects OLS estimator (due to the well known incidental parameter bias), our focus here is on improving the quality of inference by relying on the bootstrap instead of the standard normal distribution when computing critical values for test statistics. In particular, we show by simulation that confidence intervals based on the normal distribution can be very distorted in finite samples. Instead, a bootstrap that resamples the residuals and generates the bootstrap observations recursively using the estimated autoregressive panel data model greatly reduces these distortions. Thus, this method can be used to approximate the bias (as well as the entire distribution) of the (biased) fixed effects OLS estimator. In contrast to the recursive-design wild bootstrap, the fixed-design (treats the regressors as fixed when building the bootstrap data) and the pairs bootstrap (resamples observations only in the cross-section) do not consistently estimate the distribution of the standard biased fixed effects estimator and cannot be used for bias correction. This last result is crucial because Gonçalves and Kilian (2004) have established the validity of these two bootstrap procedures in pure time series autoregressive context and it implies that a naive application of these procedures to autoregressive panel data models with individual fixed effects would produce unreliable results. Another interesting finding is that the invalidity of the pairs bootstrap to estimate the distribution of the biased fixed effects estimator does not prevent this method to be valid when applied to the bias-corrected estimates.

Given the good performance of the pairwise bootstrap in the linear context, the second chapter extends its results to nonlinear dynamic panel data models with individual fixed effects. I focus on the pairs bootstrap since, unlike the recursive or fixed designed wild bootstrap considered in the first chapter, it is a non-parametric bootstrap and therefore, is generally more robust to misspecification. The usefulness of the bootstrap in the nonlinear context is crucial since nonlinearity complicates estimation and inference. As pointed out by Hahn and Newey (2004) and Hahn and Kuersteiner (2011), nonlinearity introduces an asymptotic bias in the limiting distribution of the MLE even in nonlinear static panel data models – all the regressors are strictly exogenous – in contrast to the linear case. Moreover, the MLE is generally severely biased in the nonlinear context compared to the linear context for panel data of the same sizes (n and T). My main contribution is to propose and theoretically justify the application of the pairs bootstrap in this context. I also illustrate the advantage of the bootstrap over the asymptotic theory by applying it to a canonical model of female-labor participation to construct accurate and valid confidence intervals.

In the last chapter, we consider bootstrap inference procedures for instrumental variable (IV) models with many weak instruments. It is now well known in the literature on the problem of weak instruments or weak identification that standard first-order asymptotic theory breaks down when the instruments are weakly correlated with the endogenous regressors, and commonly used IV estimators (e.g. two-stages least square (TSLS) and limited information maximum likelihood (LIML) estimators) can lose consistency; see Dufour (1997) and Staiger and Stock (1997) among others. However, as has been pointed out by Chao and Swanson (2005), having many instruments in such weakly identified situation can help to improve estimation accuracy. Indeed, using a large number of instruments can enhance the growth of the so-called concentration parameter even if each individual instrument is only weakly correlated with the endogenous explanatory variables. In this framework, Chao and Swanson (2005) have established consistency results for certain well-centered IV estimators such as the LIML estimator and Hansen, Hausman, and Newey (2008) have derived asymptotic normality results and gave Corrected Standard Errors (CSE) for these estimators. However, as shown in our simulations, CSE-based asymptotic Wald (t) tests can be very distorted in finite samples, especially when the degree of endogeneity is high. Thus, one may consider improving the quality of inference by relying on the bootstrap instead of the normal asymptotic approximation when computing critical values for test statistics. Therefore, we study the asymptotic validity of some existing bootstrap procedures for the limited information maximum likelihood (LIML) estimator when the instruments in IV regression may be weak and the number of instruments goes to infinity with the sample size. We show analytically that a standard residual-based bootstrap and the restricted efficient (RE) bootstrap of Davidson and MacKinnon (2008, 2010, 2014) cannot consistently estimate the limiting distribution of the LIML estimator. The foremost reason is that they fail to mimic well the parameter that characterizes the identification strength in the sample. Our results shed new light on bootstrap properties in the context of IV regression, highlighting in particular a fragility of bootstrap-based distributional approximations with respect to the number and the quality of instruments in the model. They also include a new, valid bootstrap-based inference procedure for IV models which is more robust to the choice of instruments, and hence exhibits demonstrably superior statistical properties over the bootstrap-based inference procedures available in the literature.

Chapter 1

Bootstrap inference for linear dynamic panel data models with individual fixed effects

1.1 Introduction

Estimation and inference in the context of linear dynamic panel data models is complicated by the presence of fixed effects. Indeed, as noted by Neyman and Scott (1948) and Nickell (1981), estimation of the fixed effects creates an incidental parameter bias in the standard fixed effects OLS estimator that persists even as $n \to \infty$ (and T is fixed). Although this inconsistency disappears when both n and T diverge to infinity, an asymptotic bias appears in the limiting distribution of the fixed effects estimator when n and T grow at the same rate, as shown by Hahn and Kuersteiner (2002). The existence of the incidental parameter bias has motivated the proposal of many bias reduction methods for panel autoregressive models with fixed effects, including Kiviet (1995), Hahn and Kuersteiner (2002), Alvarez and Arellano (2003), Bun and Carree (2005), Phillips and Sul (2007), Everaert and Pozzi (2007), Gouriéroux, Phillips, and Yu (2010), Fernandez-Val and Weidner (2013) and

Lee (2012), among others.

Our focus in this paper is on inference rather than bias correction. In particular, our main goal is to propose bootstrap methods whose finite sample properties improve upon those of the asymptotic normal approximation when computing critical values for test statistics based on bias-corrected estimators. Although this asymptotic approach is justified by the existing literature, our simulations show that asymptotic theory-based confidence intervals for the common autoregressive parameter of an AR(1) model with fixed effects can be severely distorted in finite samples. This provides motivation for the use of the bootstrap.

A natural bootstrap scheme in this context is a recursive-design residual-based bootstrap which resamples the residuals and recursively generates bootstrap observations for the dependent variable using the estimated autoregressive panel data model. The choice of how to generate the bootstrap residuals depends on the assumptions we make on the idiosyncratic error term. Here, we follow most of the existing panel data literature and maintain throughout the assumption of cross sectional independence. In contrast, we allow for time series dependence in the error term by assuming that it satisfies a martingale difference sequence assumption for each individual. This rules out serial correlation but is compatible with time series and cross sectional heteroskedasticity in the error term. To capture both forms of heteroskedasticity, we implement the residual-based bootstrap using the wild bootstrap, where bootstrap residuals are obtained by multiplying the estimated residuals by an external random variable that is i.i.d.$(0, 1)$ across both the time series and the cross sectional dimensions. A version of the recursive-design wild bootstrap method has been applied by Everaert and Pozzi (2007) for bias correction without theoretical justification.

We consider two other bootstrap methods in this paper. One is a version of the residual-based bootstrap that fixes the regressors when generating the bootstrap observations on the dependent variable (i.e. we simply add the wild bootstrap residuals to the estimated conditional mean). We call this method the fixed-design residual-based bootstrap. The other method is a pairs bootstrap which resamples the pairs formed by the dependent and the lagged dependent variables (this amounts to the standard nonparametric bootstrap applied to the pairs). Given the cross sectional independence assumption, our proposal is to resample only in the cross sectional dimension. The main reason why we also consider these two methods is that they have been applied successfully in the pure time series literature by Gonçalves and Kilian (2004), who showed that they are robust to more general forms of conditional heteroskedasticity (in the form of leverage effects) than the recursive-design residual-based bootstrap. As we will show, even though the three methods we analyze here can be viewed as panel extensions of the bootstrap methods studied by Gonçalves and Kilian (2004), the results we obtain are not a straightforward extension of the results obtained in the pure time series autoregression model due to the presence of the incidental parameter bias.

Our first finding is that only the recursive-design residual-based bootstrap is able to capture the incidental parameter bias term inherent in the fixed effects OLS estimates. The fixed-design residual bootstrap and the pairs bootstrap fail to do so as their bootstrap distributions are incorrectly centered at zero. Thus, although these bootstrap methods are more generally applicable (in that they allow for leverage effects), they do not consistently estimate the distribution of the standard fixed effects estimator in a linear dynamic panel data model with individual specific fixed effects. This is in contrast with the recursive-design bootstrap, which can be used to approximate the whole distribution of the fixed effects OLS estimator, including its bias. We formally prove the consistency of this bootstrap bias, thus providing a theoretical justification for a bootstrap based bias correction as used for instance in Everaert and Pozzi (2007).

Although our results for the recursive-design bootstrap justify bootstrap inference based on the (uncorrected and biased) fixed-effects OLS estimator without the need for an explicit bias correction, further finite sample improvements of the bootstrap approximation can be obtained if we base our inference on a bias-corrected estimator. Bootstrapping a bias-corrected fixed effects estimator essentially removes the incidental parameter bias from the asymptotic distribution, resulting in a t-statistic that is asymptotically pivotal.

Building on the theory of the bootstrap for the standard (biased) fixed effects OLS estimator, we show that the recursive-design bootstrap is asymptotically valid when applied to the bias-corrected estimator of Hahn and Kuersteiner (2002). The asymptotic invalidity of the fixed-design bootstrap for the standard fixed effects estimator extends to the bias-corrected estimator. However, as it turns out, the pairs bootstrap distribution of the bootstrap bias-corrected fixed effects estimator is consistent provided we center the bootstrap bias-corrected estimator around the bias-corrected estimator evaluated on the original sample (instead of its biased version). In the simulations, the recursive-design bootstrap is the method that does best overall, essentially removing the finite sample distortions associated with the confidence intervals that rely on the asymptotic normal distribution.

The existing literature on bootstrapping linear panel data models with fixed effects is surprisingly rather limited. One important exception is Kapetanios (2008), who proposed and studied the pairs bootstrap in the context of panel regression models with strictly exogenous regressors and fixed effects, for which the incidental parameter bias does not exist. More recently, Gonçalves (2011) proved the asymptotic validity of the moving blocks bootstrap under general forms of cross sectional and time series dependence in the regression error of a panel linear regression model. Although the regularity conditions of Gonçalves (2011) allow in principle dynamic regressors, the impact of the incidental parameter bias on inference was ruled out by assuming that $n/T \to 0$. Contrary to these papers, here we establish the consistency of the bootstrap for fixed-effects estimators when the incidental parameter bias is present. A few other papers have recently studied the validity of the bootstrap for panel data models with fixed effects and incidental parameter bias. In particular, Galvão and Kato (2013) study the asymptotic properties of the pairs bootstrap in the context of linear dynamic panel data models with possible misspecification. They find that the pairs bootstrap is asymptotically valid when applied to a bias corrected estimator and that it is robust to misspecification. Similarly, Kaffo (2013) also applies the pairs bootstrap to a bias corrected estimator in the context of nonlinear dynamic panel data models with fixed effects. In both cases, the bootstrap is not able to capture the incidental parameter bias and is only valid when used for inference on a bias corrected estimator. These results (although more general than ours) are entirely parallel to what we find here for the simpler AR(1) panel data model. However, contrary to these papers, here we are able to go a step further and propose a bootstrap method that is also able to capture the bias (the recursive-design bootstrap).

The remainder of the paper is organized as followed. Section 2 introduces the model and the assumptions, and provides a summary of the asymptotic theory for the fixed effects estimator. These results are a restatement of Hahn and Kuersteiner's (2002) results under our set of assumptions (which are slightly different from theirs). Section 3 provides the bootstrap results for the standard fixed effects OLS estimator for the three bootstrap schemes described above. We show that only the recursive-design bootstrap is able to capture the asymptotic bias term. Section 4 relies on the results of Section 3 to prove the consistency of this bootstrap method for estimating the distribution of the biased-corrected fixed effects estimator of Hahn and Kuersteiner (2002). Section 5 contains Monte Carlo results while Section 6 concludes. All proofs are relegated to the Appendix.

1.2 Assumptions and asymptotic theory for the fixed effects estimator when $n, T \to \infty$

Following Hahn and Kuersteiner (2002), we consider estimation of the autoregressive parameter θ_0 in a stationary linear dynamic panel model with fixed effects[1]

$$y_{it} = \alpha_i + \theta_0 y_{it-1} + \varepsilon_{it}, \qquad i = 1, \ldots, n; \; t = 1, \ldots, T, \tag{1.1}$$

where $|\theta_0| < 1$ and α_i are individual specific fixed effects that capture the unobserved individual heterogeneity. We assume that the initial observation y_{i0} is available. Given the stability condition that $|\theta_0| < 1$ and the assumption that the panel is stationary, the impact of initial conditions does not matter asymptotically when T is large.

The standard fixed effects OLS estimator of θ_0 is given by

$$\hat{\theta} = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right)^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right) \left(y_{it} - \bar{y}_i \right),$$

where $\bar{y}_i \equiv \frac{1}{T} \sum_{t=1}^{T} y_{it}$ and $\bar{y}_{i-} \equiv \frac{1}{T} \sum_{t=1}^{T} y_{it-1}$ are the individual time averages.

The main goal of this section is to provide a set of assumptions under which we can prove the bootstrap results that will follow and at the same time present the asymptotic theory of the fixed effects estimator under these assumptions.

Assumption A1 describes formally our set of assumptions. Note that for a given time series $\{w_t\}$ and for $j \in \mathbb{N}$, we let $cum\left(w_0, w_{t_1}, \ldots, w_{t_j-1}\right)$ denote the j^{th} order joint cumulant of $\left(w_0, w_{t_1}, \ldots, w_{t_{j-1}}\right)$ (see Brillinger, 1981, p.

[1] Our results could be generalized to higher order dynamics at the cost of complicating the notation. Since this would not add any additional insights, we prefer to follow Hahn and Kuersteiner (2002) and focus on this simple AR(1) panel model.

19), where $t_1, \ldots, t_{j-1}$ are integers[2].

Assumption A1

(i) $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ are independent across i.

(ii) For each i, $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ is a strictly stationary martingale difference sequence, i.e. $E\left(\varepsilon_{it}|\mathcal{F}_i^{t-1}\right) = 0$, a.s., where $\mathcal{F}_i^{t-1} = \sigma\left(\varepsilon_{it-1}, \varepsilon_{it-2}, \ldots\right)$.

(iii) $E\left|\varepsilon_{it}\right|^{4r}$ is uniformly bounded in i and t, for some $r \geq 2$.

(iv) $E\left(\varepsilon_{it}^2\right) = \sigma_i^2$, where $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \sigma_i^2 = \sigma^2 < \infty$.

(v) $E(\varepsilon_{it}^2 \varepsilon_{it-l} \varepsilon_{it-p}) = \tau_{ilp}$ is uniformly bounded for all i, t, $l \geq 1$, $p \geq 1$; $\tau_{ill} > 0$ for all l, and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \tau_{ilp} = \tau_{lp}$, for fixed l, $p \in \mathbb{N}$.

(vi) $\sum_{t_1,t_2,t_3=-\infty}^{+\infty} \left|cum\left(\varepsilon_{it_1}, \varepsilon_{it_2}, \varepsilon_{it_3}, \varepsilon_{i0}\right)\right| < \Delta < \infty$ uniformly in i.

(vii) $\sum_{t_1,t_2,t_3=-\infty}^{+\infty} \left|cum\left(z_{it_1}^{l_1}, z_{it_2}^{l_2}, z_{it_3}^{l_3}, z_{i0}^{l_4}\right)\right| < \Delta < \infty$ uniformly in i, l_1, l_2, l_3 and l_4, where $z_{it}^l = \varepsilon_{it}\varepsilon_{it-l}$ and $l_1,\ldots,l_4$ are positive integers.

(viii) $\frac{1}{n}\sum_{i=1}^n \left|\alpha_i\right|^2 = O(1)$.

(ix) $n, T \to \infty$ such that $n/T \to \rho < \infty$.

In this paper, we follow the fixed effects approach and treat α_i as parameters to be estimated. Accordingly, Assumption A1 implicitly treats α_i as being constant. Alternatively, our analysis can be interpreted as being conditional on a random realization of the fixed effects α_i as long as we modify our assumptions by conditioning on α_i.[3]

---

[2]In particular, $cum\left(w_0\right) = E\left(w_0\right)$ and $cum\left(w_0, w_{t_1}\right) = Cov\left(w_0, w_{t_1}\right)$. For a zero mean random variable, $cum\left(w_0, w_{t_1}, w_{t_2}\right) = E\left(w_0 w_{t_1} w_{t_2}\right)$ and $cum\left(w_0, w_{t_1}, w_{t_2}, w_{t_3}\right) = E\left(w_0 w_{t_1} w_{t_2} w_{t_3}\right) - E\left(w_0 w_{t_1}\right) E\left(w_{t_2} w_{t_3}\right) - E\left(w_0 w_{t_2}\right) E\left(w_{t_1} w_{t_3}\right) - E\left(w_0 w_{t_3}\right) E\left(w_{t_1} w_{t_2}\right)$.

[3]For instance, A1(ii) should read "For each i, $\{\varepsilon_{it}, t = 1, 2, \ldots\}$ is a strictly stationary martingale difference sequence conditional on α_i, i.e. $E\left(\varepsilon_{it}|\mathcal{F}_i^{t-1}, \alpha_i\right) = 0$, where $\mathcal{F}_i^{t-1} = \sigma\left(\varepsilon_{it-1}, \varepsilon_{it-2}, \ldots\right)$.". Similarly, all expectations should be conditional on α_i and the limits in parts (iv) and (v) should be replaced with probability limits. See Remark 1 of Hahn and Kuersteiner (2011a) for more details on the appropriate modifications.

Assumption A1(i) assumes cross sectional independence. Although we do not impose homogeneity along the cross sectional dimension, we nevertheless require this heterogeneity to disappear asymptotically. Assumption A1(ii) imposes a martingale difference sequence restriction on $\{\varepsilon_{it} : t = 1, 2, \ldots\}$ for each $i = 1, \ldots, n$; time stationarity is also assumed for simplicity. The m.d.s. assumption implies that the model for the conditional mean of y_{it} given $\mathcal{F}_i^{t-1}$ is correctly specified. This is a strong assumption that has been recently relaxed by Galvão and Kato (2013) in the context of possibly misspecified linear dynamic panel data models with fixed effects. Specifically, their results show that the pairs bootstrap is asymptotically valid for inference on a pseudo-true parameter when applied to a bias-corrected estimator. Here, we assume the model is correctly specified for the conditional mean, which allows us to obtain results for the recursive-design bootstrap based on the wild bootstrap. The motivation for this method relies on the fact that the m.d.s assumption restricts the dependence in the time dimension, ruling out serial correlation in ε_{it}, but allows for time series dependence in the form of conditional heteroskedasticity. Allowing for conditional heteroskedasticity over time is important in order to capture GARCH effects, as documented by the increasing literature on estimating large dimensional GARCH panels (see e.g. Engle, Shephard, and Sheppard (2008) and Pakel, Shephard, and Sheppard (2011)). Assumption A1(vi) restricts the fourth order cumulants of ε_{it} whereas Assumption A1(vii) is an additional eighth order restriction on the distribution of the innovations needed to establish a central limit theorem and justify covariance matrix estimation. Given that $|\theta_0| < 1$, it implies Condition 3 of Hahn and Kuersteiner (2002). Assumption A1(ix) assumes that n and T diverge to infinity at the same rate and is standard in this literature.

Under Assumption A1, we can prove the following result. See Appendix A for the proof.

Theorem 1.2.1. *Let $\{y_{it}\}$ be generated by (1.1). Under Assumption A1, we* *have*

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right) \to^d N\left(D,C\right),$$

where $D = -\sqrt{\rho}(1+\theta_0)$; *and* $C = A^{-1}BA^{-1}$, *with* $A = \sigma^2\left(1-\theta_0^2\right)^{-1}$ *and* $B = \sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\tau_{lp}$.

Theorem 1.2.1 is a restatement of Hahn and Kuesteiner's (2002) Theorem 1 under our Assumption A1. The method of proof follows closely that of Gonçalves and Kilian (2004), adapted to the panel context considered here. In particular, the cross sectional independence assumption A1(i) allows us to use results by Hansen (2007) (see also Moon and Phillips (2000) and Moon and Phillips (2004)) to derive the joint asymptotic theory of $\hat{\theta}$ as $n, T \to \infty$ under Assumption A1.

Presenting this result and its heuristic derivation is helpful in understanding the reasons for the (in)validity of the different bootstrap methods we consider in the next section. The fixed effects OLS estimator can be represented as

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right) = A_{nT}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right),$$

where $A_{nT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)^2$. Under Assumption A1, we show in the Appendix that $A_{nT} \to^P A$. Moreover, adding and subtracting $\mu_i \equiv E\left(y_{it-1}\right) = \alpha_i/\left(1-\theta_0\right)$ to the term $\left(y_{it-1}-\bar{y}_{i-}\right)$ and using the fact that the average over t of $\left(\varepsilon_{it}-\bar{\varepsilon}_i\right)$ is zero implies that

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right) = A^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right) + o_P\left(1\right).$$

The following decomposition holds for the normalized score,

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\left(\varepsilon_{it}-\bar{\varepsilon}_i\right) = \underbrace{\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\varepsilon_{it}}_{\to^d N(0,B)} \underbrace{-\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\bar{\varepsilon}_i,}_{\to^P -A\cdot D}$$

where the stochastic behavior of each of the two terms above is discussed in Lemma A.4 in Appendix A.

This result has two implications for the validity of the bootstrap. First, the bootstrap needs to mimic the asymptotic variance of $\hat{\theta}$ given by $C = A^{-1}BA^{-1}$. This variance has the usual sandwich form under conditional heteroskedasticity. In particular, it depends on the long run variance of the score process (after concentrating out α_i) defined as

$$B = \lim_{n,T\to\infty} Var\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\mu_i\right)\varepsilon_{it}\right).$$

Theorem 1.2.1 shows that B depends on[4] τ_{lp}, the limiting value of the cross sectional average of the fourth order cumulants of ε_{it}. When ε_{it} are i.i.d. $(0,\sigma^2)$, we have that $\tau_{lp} = \sigma^4$ for $l = p$ and $\tau_{lp} = 0$ for $l \neq p$, implying that $B = \sigma^4/\left(1-\theta_0^2\right)$. In this case, $B = \sigma^2 A$ and $C = 1-\theta_0^2$. But when ε_{it} are heteroskedastic (in either dimension), the fourth order cumulants of ε_{it} do not simplify and the sandwich form for C is obtained. As discussed

[4]Note that Hahn and Kuersteiner (2002) obtain a different but equivalent expression for B, given by $\frac{\sigma^4}{1-\theta_0^2}+\chi$, where $\chi \equiv \sum_{t=-\infty}^{\infty}\chi(t,0)$ and $\chi(t_1,t_2) \equiv E[u_{it_1-1}u_{it_2-1}\varepsilon_{it_1}\varepsilon_{it_2}] - E[\varepsilon_{it_1}\varepsilon_{it_2}]E[u_{it_1-1}u_{it_2-1}]$, $u_{it-1} = y_{it-1} - E\left(y_{it-1}\right)$. The constant χ reflects higher order moments of the error term when conditional heteroskedasticity is allowed for and it becomes zero when ε_{it} is i.i.d. $\left(0,\sigma^2\right)$, implying the same value for B. Our expression makes the comparison of our results with Gonçalves and Kilian (2004) easier.

by Gonçalves and Kilian (2004) in the pure time series context, bootstrap validity depends on replicating the properties of τ_{lp} and this is also true in the panel context.

Second, the bootstrap needs to capture the asymptotic bias term D created by the estimation of the fixed effects. As the decomposition above shows (and as was discussed already by Hahn and Kuersteiner (2002)), this noncentrality parameter results from the correlation between the averaged error terms $\bar{\varepsilon}_i$ and the demeaned regressors $y_{it-1} - \mu_i$ and is non zero when $\rho = \lim \frac{n}{T} \neq 0$. As we will see next, the presence of this incidental parameter asymptotic bias is the crucial difference between the application of the bootstrap in the pure time series context considered in Gonçalves and Kilian (2004) and in the panel context considered here.

1.3 Bootstrap results for the fixed effects estimator

In this section, we study the asymptotic validity of the bootstrap when applied to the fixed effects OLS estimator $\hat{\theta}$. Following Gonçalves and Kilian (2004), we consider three bootstrap methods adapted to the panel AR(1) model considered here. Two of these are residual-based wild bootstrap (WB) methods whereas the third one is a pairs bootstrap that resamples only in the cross sectional dimension (which is justified under our cross sectional independence assumption).

We use the following notation for the bootstrap asymptotics (see Chang and Park (2003a) for similar notation and for several useful bootstrap asymptotic properties): Let Z^*_{nT} be a sequence of bootstrap statistics. We write $Z^*_{nT} = o_{P^*}(1)$ in probability, or $Z^*_{nT} \to^{P^*} 0$ in probability, if for any $\varepsilon > 0$, $\delta > 0$, $\lim_{n,T\to\infty} P\left[P^*\left(|Z^*_{nT}| > \delta\right) > \varepsilon\right] = 0$. Similarly, we write $Z^*_{nT} = O_{P^*}(1)$ in probability if for all $\varepsilon > 0$ there exists a $M_\varepsilon < \infty$ such that

$$\lim_{n,T\to\infty} P\left[P^*\left(|Z_{nT}^*| > M_\varepsilon\right) > \varepsilon\right] = 0.$$

Finally, we write $Z_{nT}^* \to^{d^*} Z$ in probability if, conditional on the sample, Z_{nT}^* weakly converges to Z under P^*, for all samples contained in a set with probability converging to one. Specifically, we write $Z_{nT}^* \to^{d^*} Z$ in probability if and only if $E^*\left(f\left(Z_{nT}^*\right)\right) \to E\left(f\left(Z\right)\right)$ in probability for any bounded and uniformly continuous function f.

1.3.1 Recursive-design wild bootstrap

The recursive-design bootstrap generates a panel of pseudo observations $\{y_{it}^*,\ i = 1,\ldots,n;\ t = 1,\ldots,T\}$ recursively from the panel AR(1) model with estimated parameters,

$$y_{it}^* = \hat{\alpha}_i + \hat{\theta} y_{it-1}^* + \varepsilon_{it}^*,\ i = 1,\ldots,n;\ t = 1,\ldots,T,$$

where $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \hat{\theta} y_{it-1})$, $i = 1,\ldots,n$ and $\hat{\theta}$ is the fixed effects OLS estimator defined in the previous section (the method remains valid if $\hat{\theta}$ is replaced with any consistent estimator $\tilde{\theta}$ of θ_0). The initial condition is $y_{i0}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}}$, $i = 1,\ldots,n$, which is equivalent to setting y_{i0}^* equal to the stationary mean in the bootstrap world. The bootstrap residuals are obtained with the wild bootstrap $\varepsilon_{it}^* = \hat{\varepsilon}_{it}\eta_{it}$, where $\eta_{it} \sim$i.i.d.$(0,1)$ over (i,t) with $E^*\left|\eta_{it}\right|^4 \leq \Delta < \infty$, and $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta} y_{it-1}$ are the estimated residuals. The wild bootstrap was originally proposed by Wu (1986) and Liu (1988) in the context of cross section regressions with unconditional heteroskedasticity. Its application to the time series autoregressive context was considered by Gonçalves and Kilian (2004) (see also Kreiss (1997)). Here we extend its application to the panel autoregressive context with individual fixed effects (see Gonçalves and Perron (2014) for a recent application to panel factor models).

Letting η_{it} be i.i.d.$(0,1)$ along the two dimensions is appropriate since by Assumption A1 ε_{it} is independent across i and uncorrelated over t (due to the m.d.s. assumption), but we allow for heteroskedasticity in the two dimensions.

The bootstrap analogue of $\hat{\theta}$ is $\hat{\theta}^*_{rd}$, the recursive-design wild bootstrap OLS estimator,

$$\hat{\theta}^*_{rd} = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \bar{y}^*_{i-} \right)^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y^*_{it-1} - \bar{y}^*_{i-} \right) \left(y^*_{it} - \bar{y}^*_{i} \right), \tag{1.2}$$

where $\bar{y}^*_i$ and $\bar{y}^*_{i-}$ are defined analogously to $\bar{y}_i$ and $\bar{y}_{i-}$.

As in Gonçalves and Kilian (2004), we require a strengthening of Assumption A1 to establish the validity of the recursive-design wild bootstrap for the fixed effects OLS estimator.

A1. (v′) $\tau_{ilp} \equiv E\left(\varepsilon^2_{it}\varepsilon_{it-l}\varepsilon_{it-p}\right) = 0$ for all $l \neq p$, for all i, and t, $l \geq 1$, $p \geq 1$.

A1 (v′) is the panel analogue of Assumption A′(iv′) in Gonçalves and Kilian (2004). As they remark, this assumption further restricts the class of conditionally heteroskedastic autoregressive models that are covered by excluding certain asymmetric GARCH and ARCH models (e.g. the popular EGARCH model). This is crucial to prove that the bootstrap variance of $\hat{\theta}^*_{rd}$ is consistent for C.

Theorem 1.3.1. *Under Assumption A1 strengthened by Assumption A1(v′), it follows that*

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta}) \leq x) - P(\sqrt{nT}(\hat{\theta} - \theta_0) \leq x) \right| \to^P 0.$$

The proof of Theorem 1.3.1 is in Appendix B. The crucial difference compared to the proof of Theorem 3.2 of Gonçalves and Kilian (2004) is the need to account for the incidental parameter bias generated by the estimation of the fixed effects. In particular, Lemma .2.4 in Appendix B shows that the incidental parameter bias in the bootstrap world is such that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^{*} - \hat{\mu}_i\right) \bar{\varepsilon}_i^{*} \to^{P^*} -A \cdot D,$$

in probability, where $\hat{\mu}_i = \hat{\alpha}_i / \left(1 - \hat{\theta}\right) = E^* \left(y_{it-1}^*\right)$. This, together with the fact that

$$\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1}^{*} - \hat{\mu}_i\right) \varepsilon_{it}^{*} \to^{d^*} N\left(0, \tilde{B}\right),$$

in probability, where $\tilde{B} = \sum_{l=1}^{\infty} \theta_0^{2(l-1)} \tau_{ll}$, implies that $\sqrt{nT}\left(\hat{\theta}_{rd}^* - \hat{\theta}\right) \to^{d^*} N\left(D, A^{-1}\tilde{B}A^{-1}\right)$, in probability. Since $\tilde{B} = B$ whenever $\tau_{i,lp} = 0$ for $l \neq p$ (i.e. under A1(v′)), the recursive-design wild bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}_{rd}^* - \hat{\theta}\right)$ is consistent for the distribution of the biased fixed effects OLS estimator $\sqrt{nT}\left(\hat{\theta} - \theta\right)$. In particular, the recursive-design bootstrap contains a built-in bias correction term that mimics the incidental parameter bias induced by the individual fixed effects.

Theorem 1.3.1 justifies the construction of bootstrap percentile-type confidence intervals for θ_0 without the need for an explicit bias correction method. It does not however justify the use of the bootstrap to consistently estimate the bias of $\hat{\theta}$ without further conditions, for instance that the sequence $\left\{\sqrt{nT}\left(\hat{\theta}_{rd}^* - \hat{\theta}\right)\right\}$ is uniformly integrable (see e.g. Billingsley (1995), Theorem 25.12).

Although our focus in this paper is on using the bootstrap for constructing confidence intervals for θ_0, we now provide a result that theoretically justifies the use of the bootstrap for bias correction. The bootstrap has been used for this purpose in Everaert and Pozzi (2007) without a theoretical justification. Compared to the analytical bias correction method of Hahn and Kuersteiner (2002) (and of many others since then), the bootstrap approach is easy to generalize to more complex models without requiring the need for different analytical formulae.

Following Liu and Singh (1992) and Gonçalves and White (2005), we focus on the following bootstrap fixed effects estimator

$$\tilde{\theta}^* = \begin{cases} \hat{\theta}^*_{rd} & \text{if } \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T (y^*_{it-1} - \bar{y}^*_{i-})^2 \geq \frac{\delta}{2} \\ \hat{\theta} & \text{otherwise,} \end{cases}$$

for some $\delta > 0$. Thus, $\tilde{\theta}^*$ is equal to $\hat{\theta}^*_{rd}$ whenever $\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2$ is bounded away from zero. Since $n^{-1}T^{-1}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \to^{P^*} A > 0$, in probability, it follows that for any $\varepsilon > 0$ and sufficiently large n and T, there exists $\delta > 0$ such that

$$P\left[P^*\left(\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \geq \frac{\delta}{2}\right) > 1 - \varepsilon\right] > 1 - \varepsilon. \tag{1.3}$$

Thus, this modification does not have adverse practical consequences but at the same time it greatly simplifies the theoretical study of the bootstrap bias estimator $D^* = E^*\left(\sqrt{nT}\left(\tilde{\theta}^* - \hat{\theta}\right)\right)$.

Theorem 1.3.2. *Under the same assumptions as in Theorem 1.3.1, $D^* \to^P D$, where $D^* = E^*\left(\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right)$ and $D = -\sqrt{\rho}(1 + \theta_0)$.*

The proof of Theorem 1.3.2 is in Appendix B. We show that under Assumption A1 strengthened by A1(v′), $E^*\left(\left|\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right|^{1+\delta}\right) = O_P(1)$ for some $\delta > 0$, which is a sufficient condition for the uniform integrability of the sequence $\left\{\left|\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right|\right\}$, in probability. This together with Theorem 1.3.1 implies Theorem 1.3.2.

To end this section, we discuss bootstrap percentile-t intervals based on the following t-statistic

$$t_{\hat{\theta}^*_{rd}} = \frac{\sqrt{nT}\left(\hat{\theta}^*_{rd} - \hat{\theta}\right)}{\sqrt{\hat{C}^*_{rd}}},$$

where $\hat{C}^*_{rd} = \hat{A}^{*-1}_{rd}\hat{B}^*_{rd}\hat{A}^{*-1}_{rd}$, with

$$\hat{A}^*_{rd} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \text{ and } \hat{B}^*_{rd} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \tilde{\varepsilon}^{*2}_{it}, \tag{1.4}$$

and $\tilde{\varepsilon}^*_{it} = y^*_{it} - \bar{y}^*_i - \hat{\theta}^*\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$. The statistic $t_{\hat{\theta}^*_{rd}}$ is the bootstrap analogue of $t_{\hat{\theta}} = \sqrt{nT}\left(\hat{\theta} - \theta_0\right)/\sqrt{\hat{C}}$, where $\hat{C}$ is defined as $\hat{C}^*$ using the original data.

Given Theorems 1.2.1 and 1.3.1, the asymptotic validity of a bootstrap percentile-t interval based on $t_{\hat{\theta}^*_{rd}}$ follows from the following lemma. It shows the consistency of $\hat{C}^*_{rd}$ towards $C = A^{-1}BA^{-1}$, where $B = \tilde{B}$ under Assumption A1 (v′).

Lemma 1.3.1. *Under the same assumptions as in Theorem 1.3.1,* $\hat{C}^*_{rd} \to^{P^*} C = A^{-1}\tilde{B}A^{-1}$, *in probability.*

1.3.2 Fixed-design wild bootstrap

The fixed-design wild bootstrap generates $\{y^*_{it},\ i = 1, \dots, n;\ t = 1, \dots, T\}$ according to

$$y^*_{it} = \hat{\alpha}_i + \hat{\theta}y_{it-1} + \varepsilon^*_{it},\ i = 1, \dots, n;\ t = 1, \dots, T, \tag{1.5}$$

where $\varepsilon^*_{it} = \hat{\varepsilon}_{it}\eta_{it}$, with $\eta_{it} \sim$ i.i.d.$(0,1)$ across (i,t) such that $E^*\left|\eta_{it}\right|^4 \leq \Delta < \infty$. As for the recursive-design wild bootstrap, $\hat{\theta}$ can be replaced by any consistent estimator $\tilde{\theta}$ of θ_0 and $\hat{\alpha}_i$ by $\tilde{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \tilde{\theta}y_{it-1}\right)$.

The fixed-design wild bootstrap estimator is

$$\hat{\theta}^*_{fd} = \left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right)^2 \right)^{-1} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \bar{y}_{i-} \right) \left(y^*_{it} - \bar{y}^*_i \right). \tag{1.6}$$

Gonçalves and Kilian (2004) consider this method in the context of a pure time series autoregression and show that it is asymptotically valid for estimating the distribution of the autoregressive parameter under conditional heteroskedasticity of unknown form. In particular, and in contrast to the recursive-design wild bootstrap, the fixed-design wild bootstrap is more generally applicable because it does not require Assumption A1(v′), thus allowing for leverage effects in the form of an asymmetric response of volatility to positive and negative shocks of the same absolute magnitude. It is therefore interesting to know whether this method is valid in the context of a panel autoregression model with individual fixed effects.

Theorem 1.3.3. *Under Assumption A1, it follows that* $\sqrt{nT}\left(\hat{\theta}^*_{fd} - \hat{\theta}\right) \to^{d^*} N\left(0, C\right)$, *in probability, where* $C = A^{-1}BA^{-1}$, *with* A *and* B *defined as in Theorem 1.2.1.*

The proof of Theorem 1.3.3 is in Appendix B. In contrast to the recursive-design wild bootstrap, the fixed-design wild bootstrap is not able to reproduce the noncentrality parameter of the limiting distribution of the fixed effects OLS estimator. The bootstrap distribution of $\sqrt{nT}(\hat{\theta}^*_{fd} - \hat{\theta})$ is incorrectly centered at zero, as $n, T \to \infty$.

The reason for the failure of the fixed-design wild bootstrap to capture the incidental parameter bias is that it destroys the correlation between the average bootstrap residuals $\bar{\varepsilon}^*_i$ and the bootstrap regressors $y^*_{it-1} - \hat{\mu}_i$ because it fixes these at the sample values, i.e. $y^*_{it-1} - \hat{\mu}_i = y_{it-1} - \hat{\alpha}_i / \left(1 - \hat{\theta}\right)$. This implies that

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^{*}-\hat{\mu}_i\right)\bar{\varepsilon}_i^{*}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\hat{\mu}_i\right)\bar{\varepsilon}_i^{*}\rightarrow^{P^*}0,$$

since $E^{*}\left(\bar{\varepsilon}_i^{*}\right)=0$.

Two implications follow from this negative result. First, the fixed-design wild bootstrap cannot be used to approximate the distribution (nor the bias) of the biased fixed effects OLS estimator $\hat{\theta}$. As our simulations show, this method does not replicate the incidental parameter bias of $\hat{\theta}$ and therefore fails when used to construct percentile (or percentile-t) bootstrap confidence intervals for θ_0 based on this estimator. The second implication is that its invalidity extends to bootstrap confidence intervals for θ_0 based on the bias-corrected estimator that relies on the analytical bias correction method of Hahn and Kuersteiner (2002). We will discuss the application of the bootstrap to the bias-corrected estimator of Hahn and Kuersteiner (2002) in Section 4.

1.3.3 Pairs bootstrap

A third method that is robust to conditional heteroskedasticity of unknown form in the error term of a pure time series autoregressive model is the pairs bootstrap, where one resamples with replacement the vector that collects the dependent variable and its lagged values. This method was also studied by Gonçalves and Kilian (2004), who proved its asymptotic validity under the same assumptions as those underlying the validity of the fixed-design wild bootstrap.

The goal of this section is to study the applicability of a panel version of this bootstrap method in the context of a panel AR(1) model with individual specific fixed effects. Specifically, we consider resampling only in the cross-sectional dimension, by resampling the "pairs" (y_i, y_{i-}), where

$y_i = \begin{pmatrix} y_{i1} & \dots & y_{iT} \end{pmatrix}'$ and $y_{i-} = \begin{pmatrix} y_{i0} & \dots & y_{iT-1} \end{pmatrix}'$. This method was proposed by Kapetanios (2008) in the context of a panel regression model with strictly exogeneous regressors and fixed effects, in which case no incidental parameter bias exists[5]. Our contribution here is to analyze the properties of this method for linear dynamic panel models where the incidental parameter bias is present. Note that there are other ways of resampling the pairs (y_{it}, y_{it-1}) in the panel context. For instance, one alternative bootstrap method is to resample only in the time dimension, by resampling the "pairs" (y_t, y_{t-1}), where $y_t = \begin{pmatrix} y_{1t} & \dots & y_{nt} \end{pmatrix}'$ and $y_{t-1} = \begin{pmatrix} y_{1t-1} & \dots & y_{nt-1} \end{pmatrix}'$. This method was also considered in Kapetanios (2008) and more recently in Gonçalves (2011), who showed the asymptotic validity of the moving blocks bootstrap under general forms of cross sectional dependence and time series dependence in the regression error of a panel linear regression model. Although the regularity conditions of Gonçalves (2011) allow in principle dynamic regressors, the impact of the incidental parameter bias on inference is ruled out by assuming that $n/T \to \rho = 0$. We do not consider this bootstrap method here because we assume cross sectional independence, in which case resampling in the cross sectional dimension is more appropriate.

More specifically, we generate $\left(y_i^*, y_{i-}^*\right) \sim$ i.i.d. $\{(y_i, y_{i-1}) : i = 1, \dots, n\}$, i.e. letting $I_1, \dots, I_n$ be i.i.d. Uniform on $\{1, \dots, n\}$, we have that

$$\left(y_i^*, y_{i-}^*\right) = \begin{pmatrix} y_{I_i,1} & y_{I_i,0} \\ \vdots & \vdots \\ y_{I_i,T} & y_{I_i,T-1} \end{pmatrix}.$$

The pairs bootstrap fixed effects estimator is then defined as the original fixed effects OLS estimator but with $\{(y_{it}, y_{it-1})\}$ replaced with $\left\{\left(y_{it}^*, y_{it-1}^*\right)\right\}$. Let $\hat{\theta}_{pb}^*$ denote this estimator.

[5]See also Hounkannounon (2010) for the applicability of this method in the context of panel regression models with random effects.

Theorem 1.3.4. *Under Assumption A1, it follows that* $\sqrt{nT}\left(\hat{\theta}^*_{pb}-\hat{\theta}\right)\to^{d^*} N\left(0,C\right)$, *in probability, where* $C=A^{-1}BA^{-1}$, *with* A *and* B *defined as in Theorem 1.2.1.*

Similarly to the fixed-design wild bootstrap, the pairs bootstrap distribution of the bootstrap fixed effects OLS estimator is incorrectly centered at zero.

To understand the reason why the pairs bootstrap fails in capturing the bias, note that the pairs bootstrap fixed effects OLS estimator has the following representation

$$\sqrt{nT}\left(\hat{\theta}^*_{pb}-\hat{\theta}\right)=A^{*-1}_{nT}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)\left(\hat{\varepsilon}^*_{it}-\overline{\hat{\varepsilon}^*_{i}}\right),$$

where $A^*_{nT}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)^2$ is the bootstrap analogue of A_{nT} and $\hat{\varepsilon}^*_{it}$ is the bootstrap version of the error term ε_{it}, i.e. $\hat{\varepsilon}^*_{it}=y^*_{it}-\hat{\alpha}^*_i-\hat{\theta}y^*_{it-1}\equiv\hat{\varepsilon}_{I_i,t}$. Since ε_{it} depends on α_i (which is a function of i), its bootstrap analogue when resampling in the cross sectional dimension involves resampling $\hat{\alpha}_i$, i.e. $\hat{\varepsilon}^*_{it}$ depends on $\hat{\alpha}^*_i=\hat{\alpha}_{I_i}$, a resampled version of $\hat{\alpha}_i$. Given that resampling only occurs in the cross sectional dimension, we can define

$$s^*_i\equiv\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y^*_{it-1}-\bar{y}^*_{i-}\right)\left(\hat{\varepsilon}^*_{it}-\overline{\hat{\varepsilon}^*_{i}}\right)$$

as being the bootstrap version of $s_i\equiv\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)\left(\hat{\varepsilon}_{it}-\overline{\hat{\varepsilon}_{i}}\right)$, i.e. $s^*_i=s_{I_i}$ for all $i=1,\ldots,n$. It follows that

$$\sqrt{nT}\left(\hat{\theta}^*_{pb}-\hat{\theta}\right)=A^{*-1}_{nT}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s^*_i=A^{-1}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s^*_i}_{\to^{d^*}N(0,B)}+o_{P^*}\left(1\right),$$

given that $A^*_{nT}\to^{P^*}A$, in probability. Since $I_1,\ldots,I_n$ are i.i.d. uniformly distributed on $\{1,\ldots,n\}$, $\{s_i^* : i=1,\ldots,n\}$ is i.i.d. (conditional on the original observations) and a bootstrap CLT holds for $\frac{1}{\sqrt{n}}\sum_{i=1}^n s_i^*$, yielding an asymptotic normal distribution for $\sqrt{nT}\left(\hat{\theta}_{pb}^* - \hat{\theta}\right)$. Nevertheless, the asymptotic bootstrap population mean turns out to be zero because

$$E^*\left(s_i^*\right) = \frac{1}{n}\sum_{i=1}^n s_i = \frac{1}{n}\frac{1}{\sqrt{T}}\sum_{i=1}^n\sum_{t=1}^T \left(y_{it-1} - \bar{y}_{i-}\right)\left(\hat{\varepsilon}_{it} - \bar{\hat{\varepsilon}}_i\right) = 0,$$

by the first order condition for the fixed effects OLS estimator. Thus, the limiting bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}_{pb}^* - \hat{\theta}\right)$ is (incorrectly) centered at zero.

1.4 Bootstrapping the bias-corrected estimator

The results of Section 3 justify bootstrap inference on θ_0 based on the recursive-design bootstrap fixed effects OLS estimator $\hat{\theta}_{rd}^*$. In particular, Theorem 1.3.1 justifies the construction of bootstrap percentile intervals for θ_0 whereas Theorem 1.3.1 together with Lemma 1.3.1 justify bootstrap percentile-t intervals. Although these approaches are valid and have the advantage of avoiding the need for an explicit bias correction of $\hat{\theta}$, further finite sample improvements of the bootstrap approximation can be obtained if we base our inference on a bias-corrected estimator. Bootstrapping a bias-corrected fixed effects estimator removes the incidental parameter bias from the asymptotic distribution, resulting in a t-statistic that is asymptotically pivotal.

For the particular panel AR(1) model with individual fixed effects that we consider here, a simple analytical formula for the bias of $\hat{\theta}$ has been derived by Hahn and Kuersteiner (2002). Specifically, their bias-corrected fixed effects estimator is given by

$$\hat{\hat{\theta}} = \hat{\theta} + \frac{1}{T}\left(1 + \hat{\theta}\right), \tag{1.7}$$

where $\hat{\theta}$ is the standard biased fixed effects OLS estimator. The intuition for this bias correction is simple: by Theorem 1.2.1, $\hat{\theta} - \theta_0$ is approximately distributed as $N\left(-\frac{1}{T}(1+\theta_0), \frac{1}{nT}C\right)$. Therefore, $\Delta = -\frac{1}{T}(1+\theta_0)$ is the bias of $\hat{\theta}$ of order $O(1/T)$. The bias-corrected estimator given in (1.7) is the feasible version of the infeasible bias-corrected estimator given by $\hat{\theta} - \Delta = \hat{\theta} + \frac{1}{T}(1+\theta_0)$.

The main contribution of this section is to prove the asymptotic validity of the recursive-design bootstrap when applied to $\hat{\hat{\theta}}$. As our simulations in Section 5 show, bootstrap intervals based on $\hat{\hat{\theta}}$ have coverage probabilities that are closer to the desired nominal level than the bootstrap intervals based on $\hat{\theta}$. We also consider the application of the fixed-design and the pairs bootstrap to $\hat{\hat{\theta}}$. Our results show that whereas the asymptotic invalidity of the fixed-design bootstrap to estimate the distribution of $\hat{\theta}$ extends to $\hat{\hat{\theta}}$, this is not the case for the pairs bootstrap, which becomes a valid method of inference when used to estimate the distribution of $\hat{\hat{\theta}}$.

We start by considering the recursive-design wild bootstrap, which we now implement using only bias-corrected estimates. More specifically, the bootstrap panel observations are generated recursively from the estimated panel AR(1) model using the bias-corrected estimates, i.e. we let

$$y_{it}^* = \hat{\hat{\alpha}}_i + \hat{\hat{\theta}} y_{it-1}^* + \varepsilon_{it}^*, \; i = 1, \ldots, n; \; t = 1, \ldots, T, \tag{1.8}$$

where $\hat{\hat{\alpha}}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \hat{\hat{\theta}} y_{it-1}\right)$, $i = 1, \ldots, n$, and $\hat{\hat{\theta}}$ is the bias-corrected fixed effects OLS estimator defined in (1.7). The initial condition is $y_{i0}^* = \hat{\hat{\alpha}}_i \left(1 - \hat{\hat{\theta}}\right)^{-1}$, $i = 1, \ldots, n$.

Let $\hat{\hat{\theta}}_{rd}^*$ denote the bootstrap version of the bias-corrected fixed effects estimator (1.7), i.e.

$$\hat{\hat{\theta}}^*_{rd} = \hat{\theta}^*_{rd} + \frac{1}{T}\left(1 + \hat{\theta}^*_{rd}\right), \tag{1.9}$$

where $\hat{\theta}^*_{rd}$ is as defined in (1.2) but using bootstrap observations generated as in (1.8).

Our goal is to show the consistency of the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right)$ for the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$. An immediate consequence of Theorem 1.2.1 is that $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right) \to^d N(0, C)$ (see Theorem 2 of Hahn and Kuersteiner (2002)). Therefore, it suffices to show that $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right) \to^{d^*} N(0, C)$, in probability. This is an immediate consequence of the proof of Theorem 1.3.1. Heuristically, by replacing $\hat{\hat{\theta}}^*_{rd}$ with (1.9) we have that

$$\sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right) = \underbrace{\sqrt{nT}\left(\hat{\theta}^*_{rd} - \hat{\hat{\theta}}\right)}_{\to^{d^*} N(D,C)} + \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}^*_{rd}\right)}_{\to^{P^*} \sqrt{\rho}(1+\theta_0) \equiv -D} \quad \longrightarrow^{d^*} N(0, C),$$

where the first term converges in distribution to $N(D, C)$ by Theorem 1.3.1 (note that we center $\hat{\theta}^*_{rd}$ around $\hat{\hat{\theta}}$ because the bootstrap DGP (1.8) depends on $\hat{\hat{\theta}}$; using $\hat{\hat{\theta}}$ instead of $\hat{\theta}$ does not change the consistency result of Theorem 1.3.1 as long as we center $\hat{\hat{\theta}}^*_{rd}$ around $\hat{\hat{\theta}}$ because $\hat{\hat{\theta}}$ is a consistent estimator of θ_0). The second term converges in probability to $-D$ because $\hat{\theta}^*_{rd}$ is a consistent estimator of θ_0 (albeit biased) and $n/T \to \rho$ under Assumption A1.

Theorem 1.4.1 below states this result formally.

Theorem 1.4.1. *Under the same assumptions as in Theorem 1.3.1, we have that*

$$\sup_{x \in \mathbb{R}} \left| P^*\left(\sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right) \le x\right) - P\left(\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right) \le x\right) \right| \to^P 0.$$

Theorem 1.4.1 justifies using the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right)$ to consistently estimate the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$. The consistency of the distribution of the bootstrap t-statistic $t_{\hat{\hat{\theta}}^*_{rd}} = \sqrt{nT}\left(\hat{\hat{\theta}}^*_{rd} - \hat{\hat{\theta}}\right)/\sqrt{\hat{C}^*}$ follows whenever $\hat{C}^*$ is a consistent estimator of C, as in Lemma 1.3.1. In particular, our proposal is to choose $\tilde{C}^*_{rd} = \tilde{A}^{*-1}_{rd}\tilde{B}^*_{rd}\tilde{A}^{*-1}_{rd}$, where $\tilde{A}^*_{rd}$ and $\tilde{B}^*_{rd}$ are exactly as defined in (1.4) with the difference that $\{y^*_{it}\}$ is generated as in (1.8) and $\tilde{\varepsilon}^*_{it}$ is a function of $\hat{\hat{\theta}}^*_{rd}$ instead of $\hat{\theta}^*$. To conserve space, we do not provide the formal result but note that the same exact arguments used to prove Lemma 1.3.1 can be applied to show the consistency of $\tilde{C}^*_{rd}$ towards C. The Monte Carlo simulation results of the next section show that the finite sample properties of this approach are superior to the asymptotic normal approximation.

Next, we explain why the fixed-design bootstrap method is not asymptotically valid when applied to $\hat{\hat{\theta}}$. Let $\hat{\hat{\theta}}^*_{fd}$ denote the bootstrap version of $\hat{\hat{\theta}}$ where $\hat{\theta}^*_{fd}$ is computed as (1.6) with $\{y^*_{it}\}$ generated using equation (1.5) with $\hat{\theta}$ (and $\hat{\alpha}_i$) replaced with $\hat{\hat{\theta}}$ (and $\hat{\hat{\alpha}}_i$). Proceeding as for the recursive-design bootstrap, the following decomposition holds

$$\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right) = \underbrace{\sqrt{nT}\left(\hat{\theta}^*_{fd} - \hat{\hat{\theta}}\right)}_{\to^{d^*} N(0,C)} + \underbrace{\sqrt{\frac{n}{T}}\left(1 + \hat{\theta}^*_{fd}\right)}_{\to^{P^*} \sqrt{\rho}(1+\theta_0) \equiv -D} \quad \longrightarrow^{d^*} N\left(-D, C\right),$$

where in particular Theorem 1.3.3 justifies the convergence of the the first term. This shows that the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ is incorrectly centered at $-D$ (the correct mean should be zero since the asymptotic distribution of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$ is centered at 0).

In contrast, the pairs bootstrap is asymptotically valid when applied to $\hat{\hat{\theta}}$. In this case, letting $\hat{\hat{\theta}}^*_{pb}$ denote the bootstrap version of $\hat{\hat{\theta}}$ based on the biased fixed effects estimator $\hat{\theta}^*_{pb}$, we have that

$$\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb}-\hat{\hat{\theta}}\right)=\underbrace{\sqrt{nT}\left(\hat{\theta}^*_{pb}-\hat{\theta}\right)}_{\to^{d^*}N(0,C)}-\underbrace{\sqrt{\frac{n}{T}}\left(1+\hat{\theta}\right)}_{\to^{P}-\sqrt{\rho}(1+\theta_0)\equiv D}+\underbrace{\sqrt{\frac{n}{T}}\left(1+\hat{\theta}^*_{pb}\right)}_{\to^{P^*}\sqrt{\rho}(1+\theta_0)\equiv -D}\quad\longrightarrow^{d^*}N\left(0,C\right).$$

Thus, although the pairs bootstrap does not provide a consistent estimator of the distribution of $\sqrt{nT}\left(\hat{\theta}-\theta_0\right)$ (because its asymptotic distribution is incorrectly centered at zero), the pairs bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb}-\hat{\hat{\theta}}\right)$ is consistent for the distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}-\theta_0\right)$. The formal result is stated in the following theorem.

Theorem 1.4.2. *Under the same assumptions as in Theorem 1.3.4, we have that*

$$\sup_{x\in\mathbb{R}}\left|P^*(\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb}-\hat{\hat{\theta}}\right)\leq x)-P(\sqrt{nT}\left(\hat{\hat{\theta}}-\theta_0\right)\leq x)\right|\to^P 0.$$

For bootstrap percentile-t intervals based on the pairs bootstrap, we consider $t^*_{pb}=\sqrt{nT}\left(\hat{\hat{\theta}}^*_{pb}-\hat{\hat{\theta}}\right)/\sqrt{\tilde{C}^*_{pb}}$, with $\tilde{C}^*_{pb}=\tilde{A}^{*-1}_{pd}\tilde{B}^*_{pd}\tilde{A}^{*-1}_{pd}$, where $\tilde{A}^*_{pb}$ and $\tilde{B}^*_{pb}$ are defined as in (1.4) evaluated on the pairs bootstrap data and bias-corrected estimator. The analogue of Lemma 1.3.1 is as follows.

Lemma 1.4.1. *Under the same assumptions as in Theorem 1.3.4 ,* $\tilde{C}^*_{pb}\to^{P^*}C=A^{-1}BA^{-1}$*, in probability.*

1.5 Simulations

The goal of this section is to evaluate the finite sample performance of the three bootstrap methods studied in the previous section. We generate a panel of AR(1) processes with GARCH errors using the following equation

$$y_{it}=\alpha_i+\theta_0 y_{it-1}+\varepsilon_{it}, i=1,\ldots,n; t=1,\ldots,T, \tag{1.10}$$

where ε_{it} is such that $E\left(\varepsilon_{it}|\mathcal{F}_{it-1}\right)=0$ and $Var\left(\varepsilon_{it}|\mathcal{F}_{it-1}\right)=\sigma_{it}^2$, with

$$\sigma_{it}^2=\gamma_i\left(1-\omega-\beta\right)+\omega\varepsilon_{it-1}^2+\beta\sigma_{it-1}^2, \tag{1.11}$$

where $\gamma_i>0$, $\omega,\beta\in[0,1)$, and $\omega+\beta<1$. See Pakel, Shephard, and Sheppard (2011) for more details on this particular GARCH specification. Because $\omega+\beta<1$, these GARCH(1,1) processes are stationary but heterogeneous. In particular, the unconditional variance is given by γ_i. In the simulations, we set $\varepsilon_{it}\sim N\left(0,\sigma_{it}^2\right)$ where σ_{it} is given by (1.11) with $\sigma_{i0}^2=\gamma_i$, the unconditional variance. Following Pakel, Shephard, and Sheppard (2011), we let $\gamma_i\sim$ i.i.d. $U[0.02,0.05]$, which matches the range of annual volatility of most stock returns. The initial observations are drawn from the stationary distribution, $y_{i0}\mid\alpha_i,\gamma_i\sim N\left(\frac{\alpha_i}{1-\theta_0},\frac{\gamma_i}{1-\theta_0^2}\right)$ and we set ω and β equal to 0.30 and 0.65, respectively. Since the fixed-effects estimator is invariant to α_i, we let $\alpha_i=0$; in addition, we let $\theta_0\in\{0.3,0.6,0.9,0.99\}$, and consider $n\in\{20,\ 40,\ 60,\ 80,\ 100\}$ and $T\in\{10,\ 20,\ 30\}$.

Tables 1 and Figures 1-4 summarize our results, which are based on 2500 Monte Carlo simulations with 999 bootstrap replications each.

Table 1 reports the bias properties of the different methods. The first column corresponds to the true finite sample bias $E\left(\hat{\theta}-\theta_0\right)$ whereas the second column reports the estimated bias using the analytical correction of Hahn and Kuersteiner (2002) (i.e. $-\frac{1}{T}\left(1+\hat{\theta}\right)$). The remaining three columns pertain to the bootstrap bias estimators based on the recursive-design wild bootstrap (RD), the fixed-design wild bootstrap (FD) and the pairs bootstrap (PB). To implement the residual-based wild bootstrap methods, we let η_{it} follow the Rademacher distribution (i.e. $\eta_{it}=1$ with probability 0.5 and -1 with probability 0.5). We also used $\eta_{it}\sim N\left(0,1\right)$ and η_{it} chosen according to Mammen (1993) but these choices were dominated by the Rademacher distribution, confirming the results by Davidson and Flachaire (2008) who advocate the use of the Rademacher distribution.

The simulation results in Table 1 confirm our theory. The FD and the

Table 1.1: Performance of the bootstrap for bias-correction

					Bias		
T	n	θ_0	True	AT	RD	FD	PB
10	20	0.3	-0.098	-0.090	-0.098	0.000	0.000
		0.6	-0.138	-0.116	-0.135	0.000	-0.003
		0.9	-0.185	-0.141	-0.178	0.000	-0.005
		0.99	-0.256	-0.164	-0.232	0.000	-0.008
	60	0.3	-0.098	-0.09	-0.099	0.000	0.000
		0.6	-0.135	-0.117	-0.133	0.000	-0.001
		0.9	-0.180	-0.142	-0.174	0.000	-0.002
		0.99	-0.248	-0.165	-0.227	0.000	-0.004
	100	0.3	-0.099	-0.09	-0.099	0.000	0.000
		0.6	-0.135	-0.116	-0.133	0.000	-0.001
		0.9	-0.179	-0.142	-0.173	0.000	-0.002
		0.99	-0.245	-0.165	-0.225	0.000	-0.003
20	20	0.3	-0.049	-0.048	-0.049	0.000	0.000
		0.6	-0.070	-0.061	-0.069	0.000	-0.002
		0.9	-0.093	-0.075	-0.091	0.000	-0.004
		0.99	-0.130	-0.089	-0.124	0.000	-0.005
	60	0.3	-0.050	-0.047	-0.050	0.000	0.000
		0.6	-0.069	-0.062	-0.067	0.000	-0.001
		0.9	-0.089	-0.076	-0.088	0.000	-0.002
		0.99	-0.124	-0.089	-0.119	0.000	-0.002
	100	0.3	-0.050	-0.048	-0.050	0.000	0.000
		0.6	-0.067	-0.062	-0.067	0.000	-0.001
		0.9	-0.087	-0.076	-0.087	0.000	-0.001
		0.99	-0.122	-0.089	-0.118	0.000	-0.002

PB do not capture the incidental parameter bias whereas the RD does. An interesting result is that the RD outperforms the analytical bias correction of Hahn and Kuersteiner (2002), especially as θ_0 approaches 1.

Figures 1-4 report coverage rates of nominal 95% intervals for θ_0 based on the different bootstrap methods and the asymptotic normal distribution. We consider intervals based on $\hat{\theta}$ (Figures 1 and 2) and intervals based on its bias-corrected version $\hat{\hat{\theta}}$ (Figures 3 and 4). The bootstrap can yield both equal-tailed and symmetric intervals whereas the normal distribution generates symmetric intervals by construction. Hence, we consider bootstrap symmetric intervals in Figures 1 and 3 and bootstrap equal-tailed intervals in Figures 2 and 4. Each figure contains nine plots, where each plot shows the actual coverage rates across different values of n for a given combination of T and θ_0. Specifically, we vary T across rows ($T \in \{10, 20, 30\}$) and θ_0 across columns ($\theta_0 \in \{0.3, 0.6, 0.9\}$ for Figures 1-2 and, $\theta_0 \in \{0.6, 0.9, 0.99\}$ for Figures 3-4). All intervals are based on t-statistics studentized with an heteroskedasticity-robust standard error.

Figure 1 shows that the asymptotic theory-based intervals that rely on the biased fixed-effects estimator can be severely distorted, especially as n increases. This is entirely expected because these intervals rely on the $N(0,1)$ distribution, which does not take into account the presence of the incidental parameter bias. We only include these intervals here as a benchmark for the PB and the FD bootstrap methods, which also fail to capture this bias. The results for these methods show that they indeed follow closely the intervals based on the asymptotic standard normal distribution. Figure 1 also shows that the RD bootstrap symmetric intervals outperform all the remaining intervals, essentially eliminating the coverage distortions for $\theta_0 = 0.3$ and 0.6. For these values of θ_0, the RD bootstrap shows very little sensitivity to increases of n, which reflects the fact that it contains a built-in incidental parameter bias correction. When $\theta_0 = 0.9$, the RD bootstrap rates deteriorate (with distortions increasing as a function of n), but it still dominates

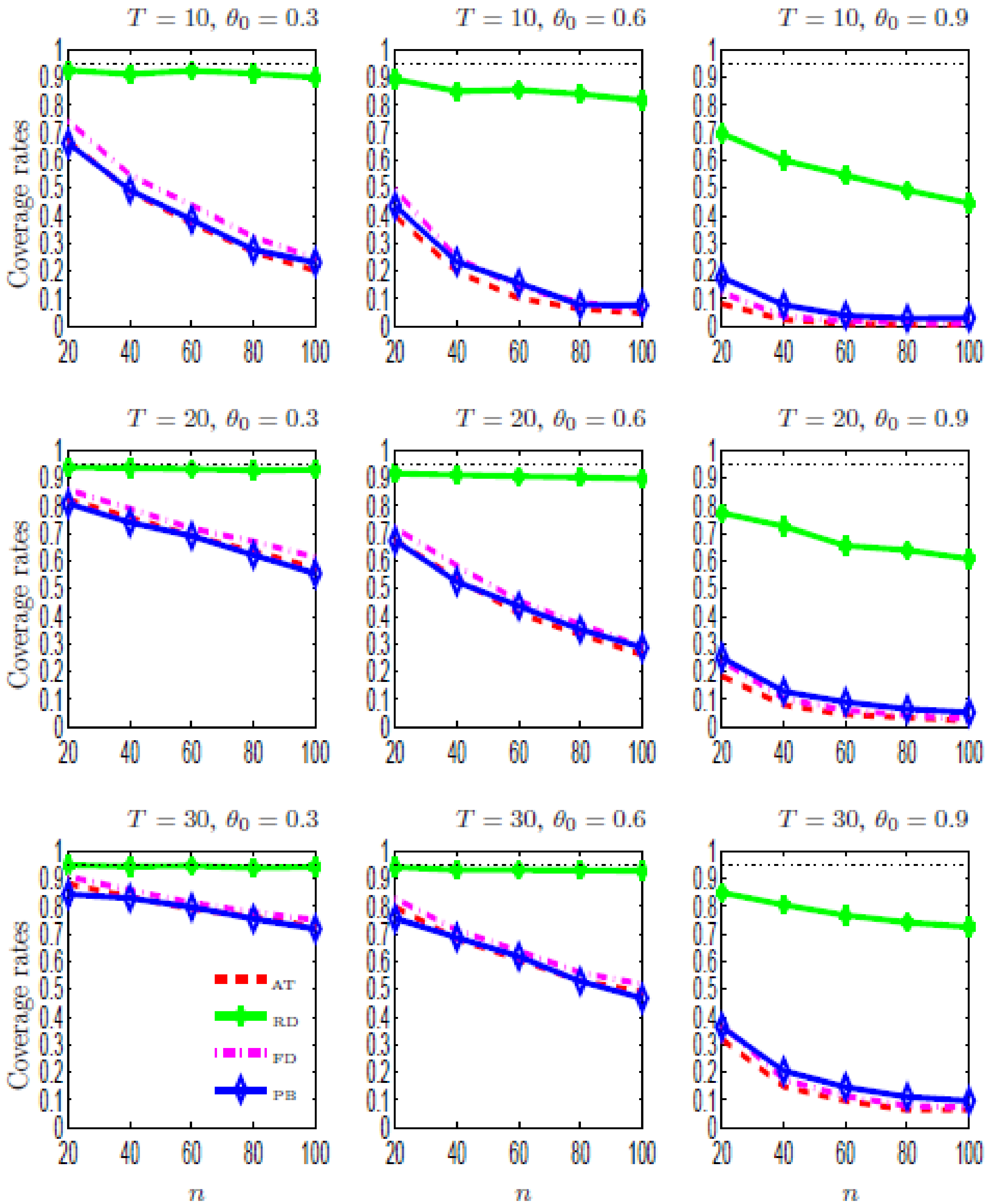

Figure 1.1: Coverage rates of nominal 95% symmetric intervals based on $\hat{\theta}$

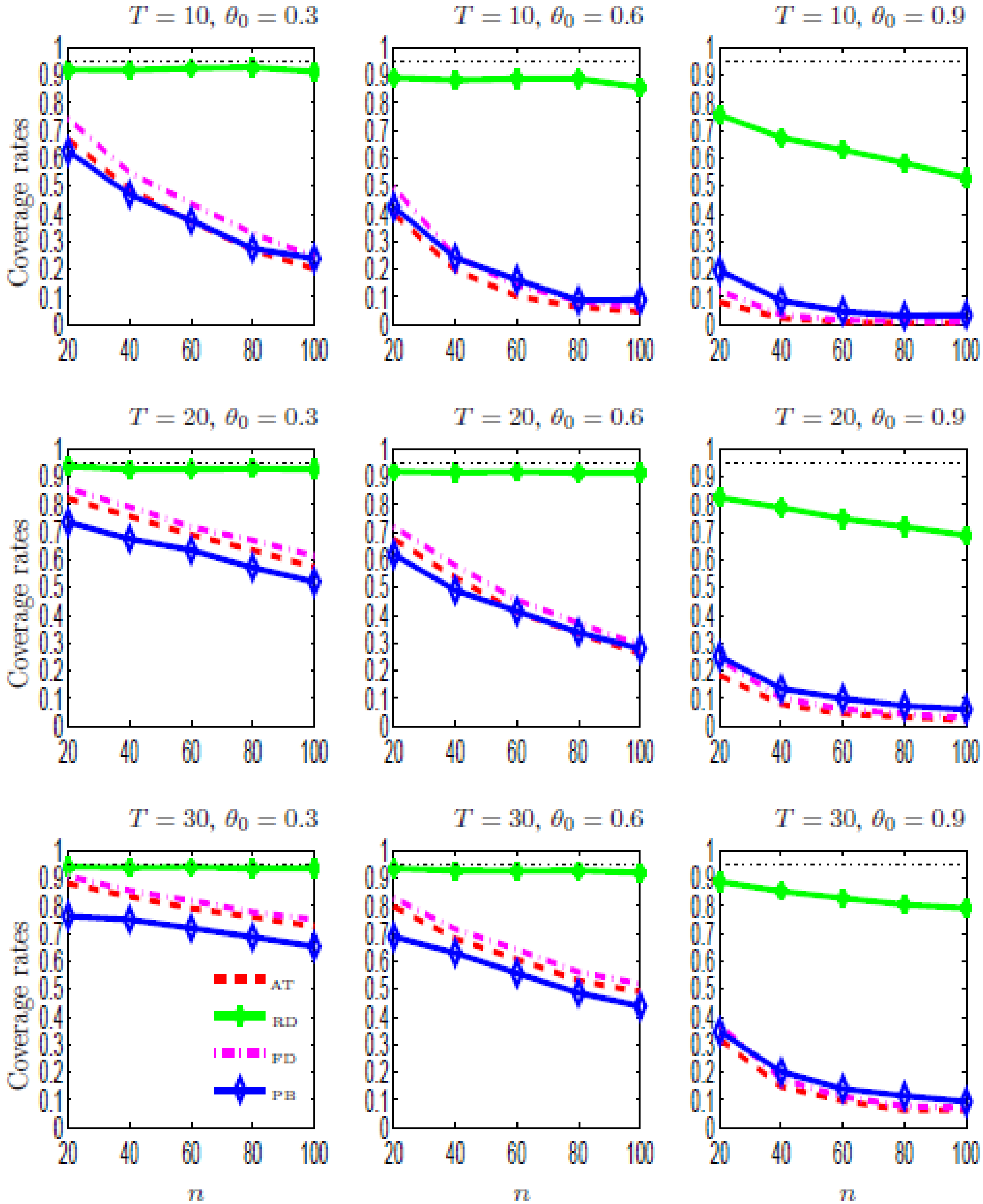

Figure 1.2: Coverage rates of nominal 95% equal-tailed confidence intervals based on $\hat{\theta}$

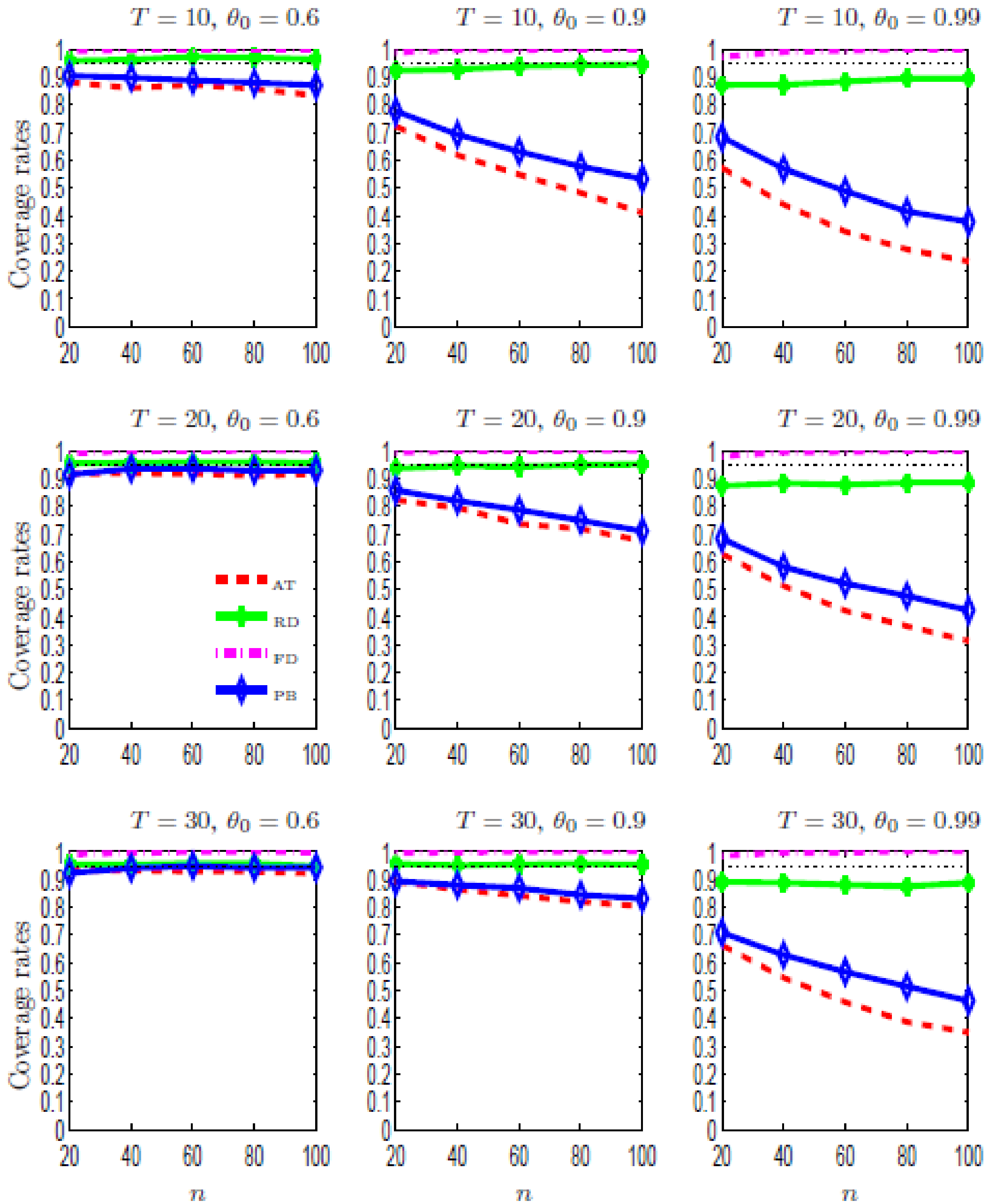

Figure 1.3: Coverage rates of nominal 95% symmetric intervals based on $\hat{\hat{\theta}}$

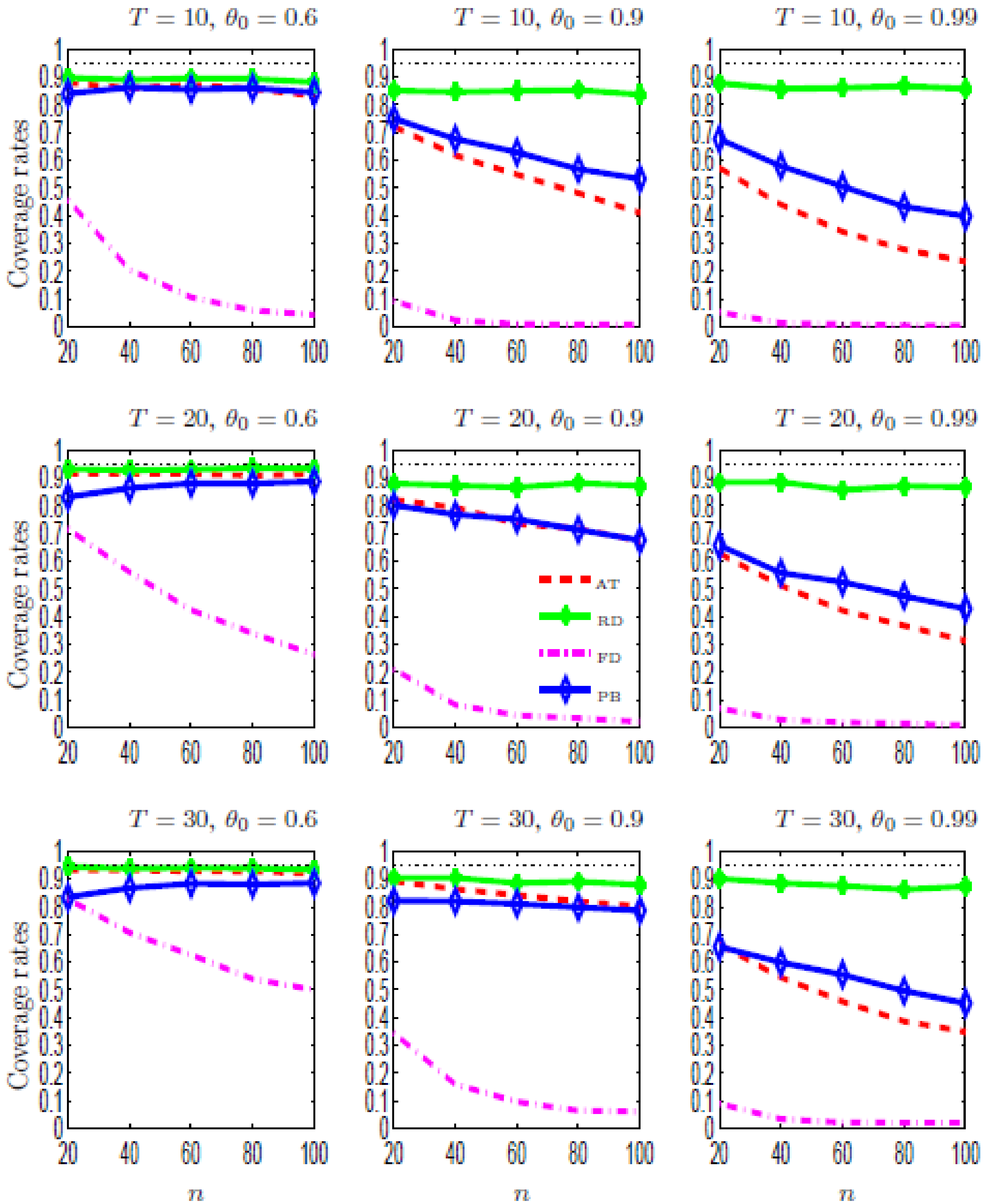

Figure 1.4: Coverage rates of nominal 95% equal-tailed confidence intervals based on $\hat{\hat{\theta}}$

the remaining methods. Distortions decrease as a function of T. As Hahn and Kuersteiner (2002) show, the limiting distribution of $\hat{\theta}$ (and its rate of convergence) changes when $\theta_0 = 1$, which explains the deterioration of all methods in the vicinity of one. The comparison of Figure 1 with Figure 2 shows that equal-tailed intervals based on the RD bootstrap outperform the symmetric intervals, especially when θ_0 is large (and close to one).

Figure 3 shows that asymptotic theory-based intervals that rely on the bias-corrected estimator $\hat{\hat{\theta}}$ can be severely distorted in finite samples, especially if θ_0 is large. In particular, large distortions arise when $\theta_0 \in \{0.9, 0.99\}$. For instance, if $T = 10$ and $\theta_0 = 0.9$, the coverage rate of a 95% asymptotic theory-based interval varies between 70% and 40% for values of n between 20 and 100. These rates increase to around 90% to 80% when $T = 30$. When $\theta_0 = 0.99$, these numbers deteriorate by a lot, varying between 70% and 35% when $T = 30$. When $\theta_0 = 0.6$, the asymptotic theory works much better, but there are still noticeable coverage distortions when $T = 10$ (rates are around 90% in this case). By comparison, the RD bootstrap symmetric intervals are much less distorted for all combinations of n, T and θ_0. For $\theta_0 \in \{0.6, 0.9\}$, this method essentially eliminates all the coverage distortions noted for the asymptotic theory-based intervals. When $\theta_0 = 0.99$, rates deteriorate but not by much, remaining around 90% for all values of n and T. The PB tends to follow the asymptotic theory-based intervals when $\theta_0 = 0.6$, but it outperforms these intervals when θ_0 increases. Symmetric intervals tend to ouperform equal-tailed intervals for these two methods, as the comparison of Figures 3 and 4 shows.

The FD bootstrap symmetric intervals are too conservative for all combinations of n, T and θ_0. The reason for this behavior is that the FD bootstrap distribution is incorrectly centered at $-D = \sqrt{\rho}\,(1 + \theta_0) > 0$. Thus, the bootstrap distribution of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ is shifted to the right of that of $\sqrt{nT}\left(\hat{\hat{\theta}} - \theta_0\right)$, implying that the bootstrap quantiles of the absolute value of $\sqrt{nT}\left(\hat{\hat{\theta}}^*_{fd} - \hat{\hat{\theta}}\right)$ will be systematically larger than those of the original finite sample distribution (centered at zero). Instead, the equal-tailed FD intervals tend to undercover, reflecting the fact that the bootstrap distribution is to the right of the true distribution. As n increases, this pushes the center of the bootstrap distribution further to the right, explaining the deterioration of the results for large values of n.

1.6 Conclusion

The main contribution of this paper is to study the validity of the bootstrap for inference on a stationary linear dynamic panel model with individual specific fixed effects. We consider three bootstrap methods: the recursive-design wild bootstrap, the fixed-design wild bootstrap and the pairs bootstrap. These methods are a natural generalization to the panel context of the bootstrap methods considered by Gonçalves and Kilian (2004) in the pure time series autoregressive model.

A crucial difference between the pure time series context and the panel context considered here is the presence of the incidental parameter bias due to the estimation of the fixed effects. We show that only the recursive-design bootstrap is able to capture this bias whereas the other two methods fail to do so. Thus, in contrast with the recursive-design wild bootstrap, the fixed-design and the pairs bootstrap do not consistently estimate the distribution of the standard biased fixed effects estimator and cannot be used for bias correction.

Although bootstrap intervals based on the biased fixed effects estimates are asymptotically valid if obtained with the recursive-design bootstrap, refinements can be obtained if bootstrap inference is based on the bias-corrected estimates. Our results show that the recursive-design is valid in this context whereas the fixed-design bootstrap is not. An interesting finding is that the invalidity of the pairs bootstrap to estimate the distribution of the biased fixed effects estimator does not prevent this method to be valid when applied to the bias-corrected estimates.

An important limitation of the present setup is the fact that we do not allow for additional regressors x_{it}.When these regressors are strictly exogeneous, a recursive-design bootstrap that fixes x_{it} at their original values should be able to capture the incidental bias. We have confirmed this by simulations (not reported here). Providing a proof of this result is outside the scope of this paper and is left for future research. The validity of the pairs bootstrap when applied to a bias-corrected estimator under the presence of extra regressors has recently been studied by Kaffo (2013) in the more general context of nonlinear dynamic models.

Further extensions of this work include the proposal of bootstrap methods that are robust to nonstationarity, where a form of the grid bootstrap can be useful, and a study of the higher order properties of the recursive-design bootstrap using Edgeworth expansions. These extensions are outside the scope of the present paper and are left for future research.

Chapter 2

Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects

2.1 Introduction

It is well known that the presence of individual fixed effects in panel data models generally causes the maximum likelihood estimator (MLE) of the parameters of interest to be inconsistent in small T large n asymptotics. Indeed, as noted by Neyman and Scott (1948) and Nickell (1981) in the linear context, estimation of the fixed effects creates an incidental parameter bias in the MLE that persists even as $n \to \infty$ (and T is fixed). Nevertheless, in large T, large n asymptotics, although the MLE is consistent, an asymptotic bias appears in its limiting distribution and T grow at the same rate, as shown by Hahn and Kuersteiner (2002). This result remains valid in the general nonlinear context. However, nonlinearity additional problems. As pointed out by Hahn and Newey (2004) and Hahn and Kuersteiner (2011b), nonlinearity introduces an asymptotic bias in the limiting distribution of the MLE even in nonlinear static panel data models – all the regressors are strictly exogenous – in contrast to the linear case. Moreover, the MLE is generally severely biased in the nonlinear context compared to the linear context for panel data of the same sizes (n and T).

The presence of the incidental parameter bias has spurred interest in bias-correction methods for nonlinear panel data models with individual fixed effects. The most prominent examples in the literature are Hahn and Kuersteiner (2011b), Arellano and Hahn (2006), Carro (2007), Fernandez-Val (2009) and Dhaene and Jochmans (2014). Except for Dhaene and Jochmans (2014), who have proposed the split-panel jackknife (SPJ) estimation, the other bias reduction methods rely on analytical corrections of either the fixed effects estimator, the moment equation or the concentrated likelihood. See Arellano and Hahn (2005) and Moon, Perron, and Phillips (2014) for an overview of the various approaches.

Our focus here is on inference rather than bias correction. Indeed, the estimation problems introduced by the individual fixed effects also affect the inference quality, since outside of automatic methods of bias correction such as the jackknife, bias reduction methods generally use in the first step a biased estimator to approximate the bias itself. This is theoretically justified in large T asymptotics since the MLE is known to be consistent, but nevertheless could lead to very imprecise estimates when T is relatively small, particularly in the nonlinear case. That could explain Dhaene and Jochmans's (2014) simulations showing that asymptotic theory-based confidence intervals for existing bias-corrected estimators can be severely distorted in finite samples. These poor finite sample performances apply also to the half-panel jackknife (HPJ) estimator. It is not surprising then that Dhaene and Jochmans (2014) have proposed the bootstrap as an alternative, but without a theoretical justification. This paper aims to provide such a theoretical justification. This is important if we want to make sure that the bootstrap helps to construct accurate and valid confidence intervals. We could consider other existing bias-corrected estimators, but here we focus on the HPJ estimator for two reasons. First, the HPJ estimator is both conceptually simple and easy to apply. Second, it has better finite sample properties than competing estimators (see Dhaene and Jochmans's (2014) Monte Carlo experiment).

Dhaene and Jochmans's (2014) bootstrap method amounts to resampling the observations only in the cross-section, which is justified under cross sectional independence. It is a natural extension of the traditional pairs bootstrap in the linear context. This traditional non-parametric bootstrap approach is known to be generally valid under mild conditions in contrast to a parametric bootstrap. Notice that we do not consider the wild bootstrap because residuals do not always exist in the type of models considered here. The asymptotic validity of the bootstrap for the HPJ estimator is established in two steps. In the first step, we show that this bootstrap method is not able to capture the incidental parameter asymptotic bias of the MLE when both n and T diverge at the same rate. It fails to do so as its bootstrap distribution is incorrectly centered at zero. Thus, this bootstrap method does not consistently estimate the distribution of the MLE for nonlinear dynamic panel data model with individual specific fixed effects. However, as we demonstrate in the second step, this method becomes asymptotically valid when used to estimate the distribution of the half-panel jackknife estimator provided we center the bootstrap HPJ estimator around the HPJ estimator evaluated on the original sample (instead of the MLE).

The existing literature on bootstrapping nonlinear panel data model from the inference perspective is surprisingly quite limited. One important exception is Sun and Kim (2013), who proposed a parametric bootstrap bias corrected maximum likelihood (ML) estimator and a double bootstrap method for inference as an alternative to asymptotic theory. Their bootstrap procedure is a version of the recursive-design bootstrap studied in Chapter 1. More recently, Gonçalves and Kaffo (2013) have studied several bootstrap methods in a linear autoregressive context and showed the validity of the cross section bootstrap when applied to the bias-corrected estimator of Hahn and Kuersteiner (2002). Using the HPJ estimator, Galvão and Kato (2013) have extended this result to linear panel data model under misspecification. The present paper extends the results of those papers to the nonlinear dynamic case.

The remainder of the paper is organized as followed. Section 2 introduces the model and the assumptions, and provides a summary of the asymptotic theory. In particular, we provide an alternative proof of the asymptotic normality of the HPJ estimator of Dhaene and Jochmans (2014) under the primitive conditions of Hahn and Kuersteiner (2011b). Section 3 provides the bootstrap results for the MLE and shows that the bootstrap is not able to capture the asymptotic bias term. Section 4, using the results of Section 3, proves the consistency of the bootstrap method for estimating the distribution of the HPJ estimator of Dhaene and Jochmans (2014). Section 5 contains Monte Carlo results. Section 6 applies the bootstrap to construct valid intervals in a canonical model of female-labor force participation and Section 7 concludes. All proofs are relegated to the Appendix.

2.2 Assumptions and asymptotic theory for the fixed effects estimator and the half-panel jackknife estimator when $n, T \to \infty$

To avoid imposing any structure on the relationship between regressors and individual heterogeneity, we follow Hahn and Kuersteiner (2011b) and adopt a fixed effects approach[1]. Suppose that we are given a panel data model with common parameter of interest θ_0 and individual fixed effects γ_{i0}, $i = 1, \ldots, n$.

[1]We treat the sample realization of the individual effects $\{\gamma_i\}_{i=1,\ldots,n}$ as parameters to be estimated.

The goal is to conduct inference on θ_0. We observe data $z_{it} = (y_{it}, x_{it})$ for units $i = 1, \ldots, n$ and time periods $t = 1, \ldots, T$. We define the maximized estimator (fixed effects estimator henceforth) of $(\theta_0, \gamma_{10}, \ldots, \gamma_{n0})$ by

$$\left(\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n\right) = \text{argmax}_{\theta,\gamma_1,\ldots,\gamma_n} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(z_{it}; \theta, \gamma_i\right),$$

where $\psi(\cdot)$ is a known criterion function that does not depend on T. In practice, ψ is often chosen to be the likelihood function for a particular parametric family of distributions. We allow x_{it} to contain lagged values of y_{it}, thus allowing for dynamic models. When the model contains lagged endogenous regressors and ψ is a likelihood function, then it is a likelihood function conditional on the lagged endogenous regressors. Thus, the estimators $\hat{\theta}, \hat{\gamma}_1, \ldots, \hat{\gamma}_n$ are the MLEs conditional on y_{i0}. To simplify notation, we assume $\dim(\gamma_i) = 1$. It is also useful to recall that the fixed effects estimator $\hat{\theta}$ is formally obtained by concentrating out the fixed effects γ_i. Letting $\hat{\gamma}_i(\theta) \equiv \text{argmax}_a \sum_{t=1}^{T} \psi(z_{it}; \theta, a)$, the fixed effects estimator of θ_0 can be rewritten as

$$\hat{\theta} = \text{argmax}_{\theta} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(z_{it}; \theta, \hat{\gamma}_i(\theta)\right).$$

As shown in the literature, when T is fixed and n large, $\hat{\gamma}_i$ is inconsistent under regularity conditions. This inconsistency carries over to $\hat{\theta}$ and the asymptotic bias of $\hat{\theta}$ is typically of order T^{-1}. This is the well known incidental parameters problem noted by Neyman and Scott (1948). However, when n and $T \to \infty$ and $n/T \to \rho < \infty$, $\hat{\theta}$ is consistent and asymptotically normal but its asymptotic distribution will be incorrectly centered when n and T grow at the same rate.

To introduce the half-panel jackknife estimator proposed by Dhaene and Jochmans (2014), suppose for simplicity that T is even, allowing for partition $\{1, \ldots, T\}$ into two half-panels, $S_1 = \{1, \ldots, T/2\}$ and $S_2 = \{T/2+1, \ldots, T\}$. If we let $\bar{\theta}_{1/2} = \frac{1}{2}\left(\hat{\theta}_{S_1} + \hat{\theta}_{S_2}\right)$ where $\hat{\theta}_{S_j}$ is the fixed effects estimator obtained from the half-panel $S_j,\ j=1,2$, then the half-panel jackknife estimator is given by

$$\hat{\theta}_{1/2}=2\hat{\theta}-\bar{\theta}_{1/2}. \tag{2.1}$$

Under high-level conditions, Dhaene and Jochmans (2014) have shown that when $n,T\to\infty$ and $n/T\to\rho<\infty$, $\hat{\theta}_{1/2}$ is asymptotically normal and its limiting distribution is free from the incidental parameters bias. However, for showing bootstrap results, we need more primitive conditions. That is why in the following, we adopt the more primitive assumptions of Hahn and Kuersteiner (2011b) and provide a new proof for this result. Let us first define:

$$\begin{aligned}
u_{it}\left(\theta,\gamma_i\right) &= \frac{\partial\psi\left(z_{it};\theta,\gamma_i\right)}{\partial\theta},\\
v_{it}\left(\theta,\gamma_i\right) &= \frac{\partial\psi\left(z_{it};\theta,\gamma_i\right)}{\partial\gamma_i},\\
U_{it}\left(\theta,\gamma_i\right) &\equiv \frac{\partial\psi\left(z_{it};\theta,\gamma_i\right)}{\partial\theta}-\rho_{i0}\frac{\partial\psi\left(z_{it};\theta,\gamma_i\right)}{\partial\gamma_i},
\end{aligned}$$

$$\begin{aligned}
\rho_{i0} &= E\left[\frac{\partial^2\psi\left(z_{it};\theta_0,\gamma_{i0}\right)}{\partial\theta\partial\gamma_i}\right]/E\left[\frac{\partial^2\psi\left(z_{it};\theta_0,\gamma_{i0}\right)}{\partial\gamma_i^2}\right],\\
\mathcal{I}_i &= -E\left[\frac{\partial U_{it}\left(\theta,\gamma_i\right)}{\partial\theta'}\right].
\end{aligned}$$

We use the short-hand notation $u_{it}\equiv u_{it}\left(\theta_0,\gamma_{i0}\right)$ and $v_{it}\equiv v_{it}\left(\theta_0,\gamma_{i0}\right)$. We will denote by $u_{it\gamma_i}$ and $u_{it\gamma_i\gamma_i}$ the first and second derivatives of u_{it} with respect to γ_i. Likewise, we will denote by $v_{it\gamma_i}$ the derivative of v_{it} with respect to γ_i.

Following Hahn and Kuersteiner (2011b), we adopt the following set of assumptions:

Assumption A

(1) For each $\eta>0$, $\inf_i\left[G_{(i)}\left(\theta_0,\gamma_{i0}\right)-\sup_{\{(\theta,\gamma):\|(\theta,\gamma)-(\theta_0,\gamma_{i0})\|>\eta\}}G_{(i)}\left(\theta,\gamma\right)\right]>$

0, where $G_{(i)}(\theta, \gamma_i) \equiv E[\psi(z_{it}; \theta, \gamma_i)]$ and $\|\cdot\|$ denotes the Euclidian norm.

(2) $n, T \to \infty$ such that $\frac{n}{T} \to \rho$ where $0 < \rho < \infty$.

(3) (i) $\{z_{it},\ t = 1, 2, \ldots,\}$ are independent across i; (ii) For each i, $\{z_{it},\ t = 1, 2, \ldots, T\}$ is a stationary mixing sequence; (iii) $\sup_i |\alpha_i(m)| \leq C a^m$ for some a such that $0 < a < 1$ and some $C > 0$, where $\mathcal{A}_t^i \equiv \sigma(z_{it}, z_{it-1}, z_{it-2}, \ldots)$, $\mathcal{B}_t^i \equiv \sigma(z_{it}, z_{it+1}, z_{it+2}, \ldots)$ and

$$\alpha_i(m) \equiv \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+m}^i} |P(A \cap B) - P(A) P(B)|.$$

(4) Let $\psi(z_{it}, \phi)$ be a function indexed by parameter $\phi = (\theta, \gamma) \in \mathrm{int}\Phi$, where Φ is a compact, convex subset of $\mathbb{R}^p$, $p = \dim(\phi) + 1$, and $R = \dim(\theta)$. Let $\nu = (\nu_1, \ldots, \nu_p)$ be a vector of nonnegative integers ν_i, $|\nu| = \sum_{j=1}^p \nu_j$ and $D^\nu \psi(z_{it}, \phi) = \partial^{|\nu|} \psi(z_{it}, \phi) / \left(\partial \phi_1^{\nu_1} \ldots \partial \phi_p^{\nu_p}\right)$. There exists a function $M(z_{it})$ such that $\|D^\nu \psi(z_{it}, \phi_1) - D^\nu \psi(z_{it}, \phi_2)\| \leq M(z_{it}) \|\phi_1 - \phi_2\|$ for all $\phi_1, \phi_2 \in \Phi$ and $|\nu| \leq 5$. The function $M(z_{it})$ satisfies $\sup_{\phi \in \Phi} \|D^\nu \psi(z_{it}, \phi)\| \leq M(z_{it})$ and

$$\sup_i E\left[|M(z_{it})|^{(10+10q)/(1-10v)+\delta}\right] < \infty$$

for some integer $q \geq p/2 + 2$, some $\delta > 0$, and $0 < v < 1/10$.

(5) Let λ_{iT} denote the smallest eigenvalue of $\Sigma_{iT} = \mathrm{Var}\left(T^{-1/2} \sum_{t=1}^T U_{it}(\theta_0, \gamma_{i0})\right)$. We assume that $\inf_i \inf_T \lambda_{iT} > 0$.

(6) $\inf_i |E[\partial v_{it}(z_{it}; \phi) / \partial \gamma_i]| > 0$ for all $\phi \in \mathrm{int}\Phi$.

(7) Let $\mu_{i1} \leq \ldots \leq \mu_{ik} \leq \ldots \leq \mu_{iR}$ be the eigenvalues of $\mathcal{I}_i$ in ascending order. Assume that (i) $0 < \inf_i \mu_{i1} \leq \sup_i \mu_{iR} < \infty$; (ii) $\lim_{n \to \infty} n^{-1} \sum_{i=1}^n \mathcal{I}_i$ exists; (iii) letting $\mathcal{I} = \lim_{n \to \infty} n^{-1} \sum_{i=1}^n \mathcal{I}_i$, we assume that $\mathcal{I}$ is positive definite.

Assumptions (1)-(5) and (7) are as in Hahn and Kuersteiner (2011b). Assumption (6) is slightly stronger than their corresponding assumption. We need it to prove the consistency of the fixed effects estimator in the bootstrap world. Hahn and Kuersteiner (2011b) provide a detailed discussion of these assumptions and show that they hold for several popular nonlinear models, including dynamic binary-choice and dynamic Tobit models with exogenous covariates.

Under these assumptions, Hahn and Kuersteiner (2011b) have shown that

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right)\to^d N\left(\beta\sqrt{\rho},\mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right),$$

where $f_i^{vU^\gamma}\equiv\sum_{l=-\infty}^{\infty}\mathrm{Cov}\left(v_{it},U_{it-l\gamma_i}\right)$, $f_i^{vv}\equiv\sum_{l=-\infty}^{\infty}\mathrm{Cov}\left(v_{it},v_{it-l}\right)$, $\varphi_i^{vU^\gamma}\equiv\lim n^{-1}\sum_{i=1}^{n}\left(E\left[v_{it\gamma_i}\right]\right)^{-1}f_i^{vU^\gamma}$, $\varphi_i^{vv}\equiv\frac{1}{2}\lim n^{-1}\sum_{i=1}^{n}\left(E\left[v_{it\gamma_i}\right]\right)^{-2}E\left[U_{it\gamma_i\gamma_i}\right]f_i^{vv}$, $\Psi=\varphi_i^{vU^\gamma}-\varphi_i^{vv}$, $\beta=-\mathcal{I}^{-1}\Psi$ and, $\Omega\equiv\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}U_{it}\right)$. $\beta\sqrt{\rho}$ is the asymptotic incidental parameters bias. The previous result follows from the following Taylor series expansion:

$$\sqrt{nT}\left(\hat{\theta}-\theta_0\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)$$
$$-\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right]\right\}+o_P(1).$$

As they pointed out, the asymptotic bias comes from the term in curly brackets. It is obvious that similar expansions will hold for the two half-panel jackknife estimators and therefore, one can show that

$$\begin{aligned}\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right) &= \sqrt{nT}\left(2\hat{\theta}-\frac{1}{2}\left(\hat{\theta}_1+\hat{\theta}_2\right)-\theta_0\right)\\ &= 2\sqrt{nT}\left(\hat{\theta}-\theta_0\right)-\frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}_1-\theta_0\right)-\frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}_2-\theta_0\right)\\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)\\ &\quad -2\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}Z_{iT}+o_P(1),\end{aligned}$$

where $Z_{iT}=Z_{1iT}+Z_{2iT}$, $T_1=T/2$ and

$$\begin{aligned}Z_{1iT} &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_1}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=T_1+1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right],\\ Z_{2iT} &= \left[\frac{1}{\sqrt{T}}\sum_{t=T_1+1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_1}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right].\end{aligned}$$

The half-panel jackknife removes the two sources for the asymptotic bias identified by Hahn and Kuersteiner (2011b). Particularly, it deletes the relevant (i) covariance of v_{it} and $U_{is}^{\gamma_i}$ and (ii) variance and autocovariance of v_{it}. We show in Appendix A that $\frac{1}{n}\sum_{i=1}^{n}Z_{iT}\to^{P}0$ and therefore, we have the following results:

Theorem 2.2.1. *Under Assumption A, we have*

$$\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right)\to^{d}N\left(0,\mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right),$$

where $\mathcal{I}\equiv\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i$ *and* $\Omega=\lim_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}Var\left(\frac{1}{\sqrt{T}}\sum_{i=1}^{T}U_{it}\right)$.

Theorem 2.2.1 corresponds to Dhaene and Jochmans's (2014) Theorem 1 applied to the half-panel jackknife estimator but now under the more primitive assumptions of Hahn and Kuersteiner (2011b). It simply states that the half-panel jackknife eliminates the bias of order T^{-1} without inflating the asymptotic variance of the fixed effects estimator. In principle, the normal approximation could be used to construct confidence intervals. However, these confidence intervals can be very size distorted in finite samples as one can see from our Monte Carlo experiment and from Dhaene and Jochmans (2014).

2.3 Bootstrap results for the fixed effects estimator

In this section, we study the asymptotic validity of the pairs bootstrap when applied to the fixed effects estimator $\hat{\theta}$. This bootstrap method amounts to constructing bootstrap samples by resampling whole cross sectional units with replacement as in Kapetanios (2008). Since we assume cross sectional independence but temporal dependence, such a resampling scheme is of great interest as it allows the application of i.i.d. bootstrap resampling rather than block bootstrap resampling. It is well known that the former enables superior approximation to distributions of statistics compared to the latter. More recently, this method was also studied by Gonçalves and Kaffo (2013). They proved that it does not capture the incidental parameter bias appearing in the limiting distribution of the fixed effects OLS estimator of the autoregressive parameter in a linear dynamic panel data models with individual fixed effects. Our contribution here is to analyze the properties of this method for nonlinear dynamic panel models. In particular, we show that, as in the autoregressive linear case, this bootstrap method does not capture the incidental parameter bias.

We assume that our bootstrap procedure resamples only in the cross-sectional dimension. More specifically, we generate $z_i^* \sim$ i.i.d. $\{z_i : i = 1, \ldots, n\}$, where $z_i^* = (z_{i1}^*, \ldots, z_{iT}^*)$ and $z_i = (z_{i1}, \ldots, z_{iT})$; i.e. letting $I_1, \ldots, I_n$ be i.i.d.

Uniform on $\{1, \ldots, n\}$, we have $z_{it}^* = z_{I_i t}$. Thus,

$$\left(\hat{\theta}^*, \hat{\gamma}_1^*, \ldots, \hat{\gamma}_n^*\right) = \text{argmax}_{\theta, \gamma_1, \ldots, \gamma_n} \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(z_{it}^*; \theta, \gamma_i\right).$$

Notice that given the resampling scheme, while $\hat{\theta}$ is the bootstrap counterpart of θ_0, $\gamma_{i0}^* \equiv \hat{\gamma}_{I_i}$ is the bootstrap counterpart of γ_{i0}. Let $\hat{\gamma}_i^*(\theta) \equiv \text{argmax}_a \sum_{t=1}^{T} \psi\left(z_{it}^*; \theta, a\right)$. It is also obvious that, $\hat{\gamma}_i^*(\theta) = \hat{\gamma}_{I_i}(\theta)$ for all θ where $\hat{\gamma}_i(\theta) \equiv \text{argmax}_a \sum_{t=1}^{T} \psi\left(z_{it}; \theta, a\right)$. Therefore, $\hat{\theta}^*$ can be rewritten as

$$\hat{\theta}^* = \text{argmax}_\theta \sum_{i=1}^{n} \sum_{t=1}^{T} \psi\left(z_{it}^*; \theta, \hat{\gamma}_i^*(\theta)\right).$$

We also introduce the following bootstrap notations:

$$\begin{aligned} u_{it}^*(\theta, \gamma_i) &= \frac{\partial \psi\left(z_{it}^*; \theta, \gamma_i\right)}{\partial \theta}, \\ v_{it}^*(\theta, \gamma_i) &= \frac{\partial \psi\left(z_{it}^*; \theta, \gamma_i\right)}{\partial \gamma_i}. \end{aligned}$$

Our first main result is given in Theorem 2.3.1. It simply states that the pairs bootstrap does not capture the incidental parameters bias.

Theorem 2.3.1. *Under Assumption A, we have*

$$\sqrt{nT}\left(\hat{\theta}^* - \hat{\theta}\right) \to^{d^*} N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right) \text{ in probability}$$

with $\mathcal{I}$ and Ω defined as in Theorem 2.2.1.

To understand the reason why the pairs bootstrap fails in capturing the bias, note that the pairs bootstrap fixed effects estimator has the following representation

$$\sqrt{nT}\left(\hat{\theta}^* - \hat{\theta}\right) = \mathcal{I}_{nT}^{*-1} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it}^*\left(\hat{\theta}, \gamma_{i0}^*\right), \tag{2.2}$$

where

$$\mathcal{I}_{nT}^{*}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)-\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}v_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)\right\}$$

is the bootstrap counterpart of $\mathcal{I}_n \equiv \frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i$ and $\tilde{\theta}^*$ lies between $\hat{\theta}^*$ and $\hat{\theta}$. Given that resampling only occurs in the cross sectional dimension and also that $\gamma_{i0}^* \equiv \hat{\gamma}_{I_i}$, we can define

$$s_i^* \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right)$$

as being the bootstrap version of $s_i \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)$, i.e. $s_i^* = s_{I_i}$ for all $i = 1,\ldots,n$. It follows that

$$\sqrt{nT}\left(\hat{\theta}^*-\hat{\theta}\right)=\mathcal{I}_{nT}^{*-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^*=\mathcal{I}^{-1}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^*}_{\to^{d^*}N(0,\mathcal{I}^{-1}\Omega\mathcal{I}^{-1})}+o_{P^*}(1),$$

given that $\mathcal{I}_{nT}^* \to^{P^*} \mathcal{I}$, in probability. Since $I_1,\ldots,I_n$ are i.i.d. uniformly distributed on $\{1,\ldots,n\}$, $\{s_i^* : i = 1,\ldots,n\}$ is i.i.d. (conditional on the original observations) and a bootstrap CLT holds for $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s_i^*$, yielding an asymptotic normal distribution for $\sqrt{nT}\left(\hat{\theta}^*-\hat{\theta}\right)$. Nevertheless, the asymptotic bootstrap population mean turns out to be zero because

$$E^*\left(s_i^*\right)=\frac{1}{n}\sum_{i=1}^{n}s_i=\frac{1}{n}\frac{1}{\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)=0,$$

by the first order condition for the fixed effects estimator. Thus, the limiting bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}^*-\hat{\theta}\right)$ is (incorrectly) centered at zero.

2.4 Bootstrapping the half-panel jackknife estimator

The results of Section 2.3 show that bootstrap inference on θ_0 based on the pairs bootstrap fixed effects estimator $\hat{\theta}^*$ is not valid. However, as in the linear autoregressive case (see chapter 1), this bootstrap method becomes valid when applied to the half-panel jackknife estimator. The main contribution of this section is to prove the asymptotic validity of the pairs bootstrap when applied to $\hat{\theta}_{1/2}$. We assume again that T is even and partition $\{1,\ldots,T\}$ into two half-panels, $S_1 = \{1,\ldots,T/2\}$ and $S_2 = \{T/2+1,\ldots,T\}$. If we let $\bar{\theta}^*_{1/2} = \frac{1}{2}\left(\hat{\theta}^*_{S_1} + \hat{\theta}^*_{S_2}\right)$ where $\hat{\theta}^*_{S_j}$ is the fixed effects estimator obtained from the bootstrap half-panel S_j, $j = 1,2$, then the bootstrap half-panel jackknife estimator is given by

$$\hat{\theta}^*_{1/2} = 2\hat{\theta}^* - \bar{\theta}^*_{1/2}. \tag{2.3}$$

Our goal is to show the consistency of the bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}^*_{1/2} - \hat{\theta}_{1/2}\right)$ for the distribution of $\sqrt{nT}\left(\hat{\theta}_{1/2} - \theta_0\right)$. Therefore, it suffices to show that $\sqrt{nT}\left(\hat{\theta}^*_{1/2} - \hat{\theta}\right) \to^{d^*} N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right)$, in probability. This is an immediate consequence of (2.2) and the proof of Theorem 2.3.1. Heuristically, by replacing $\hat{\theta}^*_{1/2}$ and $\hat{\theta}_{1/2}$ with (2.3) and (2.1) respectively, we have that

$$\begin{aligned}
\sqrt{nT}\left(\hat{\theta}^*_{1/2} - \hat{\theta}_{1/2}\right) &= 2\sqrt{nT}\left(\hat{\theta}^* - \hat{\theta}\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_1 - \hat{\theta}_1\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_2 - \hat{\theta}_2\right) \\
&= 2\mathcal{I}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) - \frac{\sqrt{2}}{2}\mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=1}^{T_1} u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) \\
&\quad -\frac{\sqrt{2}}{2}\mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T} u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) + o_{P^*}(1) \\
&= \mathcal{I}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) + o_{P^*}(1) \to^{d^*} N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right),
\end{aligned}$$

where $T_1 = T/2$ and the distributional result in the last equality follows from the proof of Theorem 2.3.1. Thus, although the pairs bootstrap does not provide a consistent estimator of the distribution of $\sqrt{nT}\left(\hat{\theta}-\theta_0\right)$ (because its asymptotic distribution is incorrectly centered at zero), the pairs bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}^*_{1/2}-\hat{\theta}_{1/2}\right)$ is consistent for the distribution of $\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right)$. The formal result is stated in the following theorem.

Theorem 2.4.1. *Under the same assumptions as in Theorem 2.3.1, we have that*

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\sqrt{nT}\left(\hat{\theta}^*_{1/2}-\hat{\theta}_{1/2}\right)\leq x\right)-P\left(\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right)\leq x\right)\right|\to^P 0.$$

Theorem 2.4.1 justifies using the bootstrap distribution of $\sqrt{nT}\left(\hat{\theta}^*_{1/2}-\hat{\theta}_{1/2}\right)$ to consistently estimate the distribution of $\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right)$.

2.5 Simulations

In order to evaluate the bootstrap's finite sample performance, we conduct some Monte Carlo experiments. We focus on the following standard dynamic binary-choice model:

$$y_{it} = 1\left\{\alpha_{i0}+\rho_0 y_{it-1}+\beta_0 x_{it}\geq u_{it}\right\},\ t=1,\ldots,T-1,\ i=1,\ldots,n,$$

where $u_{it}\sim$ i.i.d. $\mathcal{N}(0,1)$. We consider two data generating processes (DGPs) for the strictly exogenous regressor x_{it} and the individual fixed effects α_{i0} . The first DGP is similar to Hahn and Kuersteiner (2011b) probit design with i.i.d. exogenous regressor. We generate $x_{it}\sim$ i.i.d. $\mathcal{N}(0,1)$, $\alpha_{i0}=3\left(x_{i0}+x_{i1}+x_{i2}+x_{i4}\right)/4$ and $y_{i0}=1\left\{\alpha_{i0}+\beta_0 x_{i0}\geq u_{i0}\right\}$. In the second DGP, we generate $\alpha_{i0}\sim$ i.i.d. $\mathcal{N}(0,1)$ and

$$x_{it} = .5\alpha_{i0}+.5x_{it-1}+v_{it},\ t=1,\ldots,T-1,\ i=1,\ldots,n.$$

As in Dhaene and Jochmans (2014) Monte Carlo experiment, y_{i0} and x_{i0} are drawn from their respective stationary distributions. Notice that in both DGPs, the unobserved heterogeneity is correlated with the exogenous variable. We set $n = 500$; $T = 6, 9, 12, 15$; $\rho_0 = .5, 1$; $\beta_0 = .5$; and ran 1,000 Monte Carlo replications and 399 bootstrap draws.

Table 1 and 2 report coverage rates of nominal 95% intervals for (ρ_0, β_0) based on the bootstrap and the asymptotic normal distribution. We consider both intervals based on the MLE $\left(\hat{\rho}, \hat{\beta}\right)$ and intervals based on the HPJ estimator $\left(\hat{\rho}_{1/2}, \hat{\beta}_{1/2}\right)$. The first 4 rows of each table pertains to $(\rho_0, \beta_0) = (.5, .5)$ while the four last rows pertain to $(\rho_0, \beta_0) = (1, .5)$. Asymptotic theory-based confidence intervals are based on plug-in estimates of the asymptotic variance, computed from the profile Hessian whereas bootstrap confidence intervals are computed using both percentile and percentile-t methods.

Table 2.1: Probit model with i.i.d. regressors

		$\hat{\rho}$			$\hat{\rho}_{1/2}$			$\hat{\beta}$			$\hat{\beta}_{1/2}$		
T	ρ_0	Asy	Perc	Perc-t	Asy	Perc	Perc-t	Asy	Perc	Perc-t	Asy	Perc	Perc-t
6	0.5	0.0	0.0	0.0	18.7	48.0	48.2	21.0	30.6	11.1	53.0	79.6	80.0
9	0.5	0.0	0.0	0.0	82.6	91.4	91.5	30.8	39.1	20.3	85.4	90.5	90.4
12	0.5	0.1	0.1	0.0	90.1	94.5	94.5	43.1	46.7	31.9	92.6	94.6	94.5
15	0.5	0.5	0.5	0.2	92.1	93.8	94.0	46.1	50.2	35.7	94.2	95.2	95.2
6	1	0.0	0.0	0.0	39.9	70.6	70.1	24.4	32.9	10.4	64.1	87.9	88.4
9	1	0.0	0.0	0.0	88.1	95.3	95.0	34.0	41.5	20.0	87.1	93.4	93.4
12	1	0.0	0.0	0.0	90.5	94.6	94.6	44.9	50.9	27.8	92.8	94.6	94.4
15	1	0.4	0.4	0.1	92.9	95.4	95.3	46.3	50.5	32.5	94.4	94.8	95.3

As predicted by the theory, asymptotic theory- and bootstrap-based confidence intervals that rely on the MLE are severely distorted. In additional simulations (not reported here), we found that the bootstrap is not able to capture the bias of the MLE if one uses the traditional bootstrap bias correction formula $E^* \left(\hat{\theta}^* - \hat{\theta}\right)$ to approximate $E \left(\hat{\theta} - \theta_0\right)$. Indeed, $E^* \left(\hat{\theta}^* - \hat{\theta}\right)$ was close to zero for both models, implying that the bootstrap distribution is incorrectly centered at zero. This explains why the performance of bootstrap intervals based on the MLE is similar to that of asymptotic theory-based intervals since the later does not also take into account the presence of the incidental parameter bias.

Table 2.2: Probit model with AR(1) regressor

		$\hat{\rho}$			$\hat{\rho}_{1/2}$			$\hat{\beta}$			$\hat{\beta}_{1/2}$		
T	ρ_0	Asy	Perc	Perc-t	Asy	Perc	Perc-t	Asy	Perc	Perc-t	Asy	Perc	Perc-t
6	0.5	0.0	0.0	0.0	22.7	54.4	54.4	5.3	9.7	3.2	58.0	90.1	90.4
9	0.5	0.0	0.0	0.0	79.9	90.2	89.8	7.1	13.2	4.6	82.3	94.5	94.9
12	0.5	0.0	0.0	0.0	91.0	95.0	95.0	11.1	15.0	7.8	90.1	95.9	96.1
15	0.5	0.3	0.3	0.0	92.1	95.8	95.8	16.0	20.6	11.9	90.6	95.3	95.4
6	1	0.0	0.0	0.0	48.6	78.7	78.2	6.5	11.2	2.7	67.3	94.9	95.0
9	1	0.0	0.0	0.0	86.1	93.3	93.1	8.8	14.6	4.5	83.9	95.5	95.6
12	1	0.0	0.1	0.0	91.3	94.4	94.5	11.0	16.4	6.9	88.6	95.2	95.4
15	1	0.2	0.2	0.0	91.2	94.0	94.0	16.0	21.5	9.9	90.7	96.1	96.0

On the other hand, we see that the confidence intervals based on the HPJ estimator have much improved coverage compared to the MLE, especially when computed by bootstrapping. The advantage of the bootstrap on the asymptotic theory is more important when T is small. For instance, if $T = 9$, the coverage rate of a 95% bootstrap-based confidence intervals of θ_0 are approximately 10% higher than the coverage of asymptotic theory-based intervals for both DGPs. The poor performances of both asymptotic theory- and bootstrap-based confidence intervals for the HPJ estimator when T is small ($T = 6$) may be explained by the fact the limiting distribution of the HPJ estimator is derived under the asymptotics where T grows to infinity at the same rate as n.

2.6 Empirical application: Female labor-force participation

This section applies our bootstrap method to the inter-temporal labor force participation of married women and allows us to construct valid confidence intervals. Female labor-force participation may prove especially important in the years to come, as an ageing population will place an increasingly severe burden on public finances of developed countries. Old age pension expenditure will increase, as will government outlays for health care. Low birth rates will add to the problem, and a shrinking working-age population will have to provide for an increasing number of pensioners. Higher labor force participation and longer careers are important parts of the solution. In addition to reforming pension schemes, many countries now see a need to make use of the large unused work potential among women.

The data used in the analysis pertain to the nine calendar years 1979-1988 of the Panel Study of Income Dynamics (PSID), corresponding to waves 13-22 of the PSID. The sample consists of 1461 continuously married couples, aged between 18 and 60 in 1985, and the husband is a labor force participant in each of the sample years. During the sampling period, 664 women changed participation status at least once. Following Dhaene and Jochmans (2014), we consider a fixed effect approach. The empirical specification for modeling intertemporal participation decisions involves the following dynamic reduced form specification:

$$y_{it} = 1\left\{\alpha_{i0} + \rho_0 y_{it-1} + \beta_0 x_{it} \geq u_{it}\right\}, \; u_{it} \text{ i.i.d. } \mathcal{N}(0,1),$$

where y_{it} is an indicator for labor-force participation of individual i at time t and x_{it} is a vector of time-varying covariates. The time-varying covariates include the number of children of at most two years of age (# children 0-2), between 3 and 5 years of age (# children 3-5), and between 6 and 17 years of age (# children 6-17), as well as the log of the husband's earnings (log husband income), a quadratic function of age, and a set of year dummies. This setup coincides with that of Dhaene and Jochmans (2014), Fernandez-Val (2009) and is similar to that of Carro (2007).

Table 3 shows the estimation results by MLE and HPJ estimator with the associated asymptotic and bootstrap confidence intervals; 999 bootstrap replications were used. In line with the literature, there are significant differ-

Table 2.3: Female labor-force participation: fixed-effect probit estimates and confidence intervals

	Dependant variable: participation							
	Regressor	Estimates	Asymptotic theory		Percentile		Percentile-t	
MLE	Lagged participation	0.757	[0.672	0.841]	[0.667	0.846]	[0.667	0.847]
	# children 0-2	-0.560	[-0.674	-0.445]	[-0.730	-0.390]	[-0.716	-0.403]
	# children 3-5	-0.300	[-0.407	-0.194]	[-0.445	-0.155]	[-0.442	-0.158]
	# children 6-17	-0.091	[-0.176	-0.006]	[-0.202	0.020]	[-0.201	0.019]
	log husband income	-0.241	[-0.350	-0.131]	[-0.370	-0.112]	[-0.368	-0.114]
HPJ	Lagged participation	1.322	[1.229	1.414]	[1.204	1.440]	[1.201	1.443]
	# children 0-2	-0.720	[-0.845	-0.596]	[-0.916	-0.525]	[-0.911	-0.529]
	# children 3-5	-0.414	[-0.531	-0.296]	[-0.642	-0.185]	[-0.631	-0.196]
	# children 6-17	-0.119	[-0.212	-0.026]	[-0.318	0.081]	[-0.321	0.083]
	log husband income	-0.284	[-0.399	-0.169]	[-0.453	-0.115]	[-0.451	-0.117]

ences among the estimated parameters. The MLE underestimates the effect of the lagged participation while the adjustment to the other coefficients is smaller. Although bootstrap-based confidence intervals are always larger than asymptotic theory-based confidence intervals, both types of intervals are very similar for the MLE but significantly different for the HPJ estimator. In the case of MLE, it confirms our theoretical results since both intervals rely on critical values of distributions incorrectly centered at zero. The fact that bootstrap intervals based on the HPJ estimator are significantly larger than their asymptotic theory counterparts is also in line with our simulation findings, suggesting that the latter will generally undercover compared to the former, especially when T is relatively small (e.g. $T = 9$). Basically, asymptotic theory-based confidence intervals will underestimate the uncertainty about the parameters. This last result illustrates the superiority of the bootstrap compared to the asymptotic theory, even after applying a bias correction method.

Interestingly, the coefficient on the number of children between 6 and 17 years of age is significantly different from zero according to the asymptotic theory (zero is not included in the asymptotic theory-based confidence interval) while the bootstrap concludes that it is not the case (zero is included in the bootstrap confidence interval). This result underscores the importance of having an accurate confidence interval. It means that, ceteris paribus, increasing the number of children of at least 6 years old does not affect women's labor participation. In contrast to the prediction of the asymptotic theory, this last result is intuitive given that children generally start to go to school at 6 years of age and thus, women are less time constrained and have more free time for searching and keeping a job. From the policymaker point of view, it implies that in designing incentive measures to encourage female labor participation, focusing on women with children of at most 5 years of age – rather than all women – was enough.

2.7 Conclusion

In this paper we studied the validity of the bootstrap for inference in a stationary nonlinear dynamic panel model with individual specific fixed effects. Proposed by Dhaene and Jochmans (2014) as an alternative to the asymptotic theory, this bootstrap method amounts to resampling of the observations only from the cross-section. We show that this bootstrap method is not able to capture the incidental parameter asymptotic bias of the MLE when both n and T are large. It fails to do so as its bootstrap distribution is incorrectly centered at zero. Thus, this bootstrap method do not consistently estimate the distribution of the MLE for nonlinear dynamic panel data model with individual specific fixed effects. However, an interesting finding is that the invalidity of the pairs bootstrap to estimate the distribution of the biased MLE does not prevent this method to be valid when applied to the HPJ estimator. In Monte Carlo experiments, bootstrap confidence intervals that rely on the HPJ estimator perform well in relatively short panels with much improved coverage relative to asymptotic theory-based intervals. Questions for future research include the proposal of bootstrap methods that are robust to nonstationarity and/or misspecification of the likelihood.

Chapter 3

Bootstrap Inference for Instrumental Variable Models with Many Weak Instruments

3.1 Introduction

Empirical applications of instrumental variables (IV) estimation often produce imprecise results. It is now well known in the literature on the problem of weak instruments or weak identification that standard first-order asymptotic theory breaks down when the instruments are weakly correlated with the endogenous regressors, and commonly used IV estimators (e.g. two-stage least squares (TSLS) and limited information maximum likelihood (LIML) estimators) can lose consistency; see Dufour (1997) and Staiger and Stock (1997) among others. However, as has been demonstrated by Chao and Swanson (2005), having many instruments in such weakly identified situation can help to improve estimation accuracy. Indeed, using a large number of instruments can enhance the growth of the so-called concentration parameter even if each individual instrument is only weakly correlated with the endogenous explanatory variables. Chao and Swanson (2005) show that for certain well-centered IV estimators such as LIML, consistency can be established even when instrument weakness is such that the rate of growth of the concentration parameter, r_n, is slower than l, the number of instruments, and possibly much slower than the sample size n, provided that $\sqrt{l}/r_n \to 0$ as $n \to \infty$. Hansen, Hausman, and Newey (2008) reveal in an application from Angrist and Krueger (1991) that using 180 instruments, rather than 3, substantially improves estimator accuracy.

Moreover, for implementing Wald-type inferences in the context of many weak instruments, Hansen, Hausman, and Newey (2008) derive asymptotic normality of LIML and its modified version proposed by Fuller (1977), and provide Corrected Standard Errors (CSE) for these estimators. The CSE are an extension of those of Bekker (1994) and are asymptotically correct under a variety of frameworks proposed in the IV literature, including the many weak instrument sequence of Chao and Swanson (2005) and Stock and Yogo (2005), as well as the many instrument sequence of Kunitomo (1980), Morimune (1983) and Bekker (1994). Recently, the CSE have been further extended by Chao, Swanson, Hausman, Newey, and Woutersen (2012), Hausman, Newey, Woutersen, Chao, and Swanson (2012) to the heteroskedastic case and by Newey and Windmeijer (2009) to continuously updating GMM (CUE) and other generalized empirical likelihood (GEL) estimators.

However, our simulation evidence shows that CSE-based asymptotic Wald (t) tests can be significantly distorted in finite samples, especially in the case of strong endogeneity. This provides motivation for the use of bootstrap tests instead of the CSE-based asymptotic Wald (t) tests to improve the quality of inference. Furthermore, the CSE have a rather tedious form and thus can be difficult to implement in practice; this also motivates the use of bootstrap methods. In particular, the bootstrap would help to avoid computing the tedious form of the CSE if percentile type bootstrap approximations are valid under many/many weak instruments.

The existing literature on bootstrapping IV models turns out to be rather limited. Moreira, Porter, and Suarez (2009) provide theoretical proof that guarantees the bootstrap validity of Kleibergen (2002)'s score statistic even under Staiger and Stock (1997)'s weak instrument asymptotics, in which the coefficients of the instruments are specified to be in an $n^{-1/2}$ shrinking neighborhood of zero and the number of instruments is kept fixed. Davidson and MacKinnon (2008, 2010) study various bootstrap procedures (pairs bootstrap and residual-based bootstrap) for hypothesis testing in the IV model. Their extensive simulation results show that the bootstrap approaches typically perform very well relative to the normal approximation, including the case in which instruments are quite weak. Moreover, in their recent paper, Davidson and MacKinnon (2014) study asymptotic and bootstrap methods of constructing confidence sets in a similar context. Their simulation evidence reveals that bootstrap confidence sets obtained by inverting the t-statistic based on LIML estimates perform very well relative to confidence sets using asymptotic critical values. However, all these papers focus on the case where the number of instruments is kept small relative to the sample size.

In this paper, we study bootstrap-based inference methods under many/many weak instrument sequences. Based on the impressive results for the models with a small number of instruments, one may expect the bootstrap to also perform well when the number of instruments becomes large. Surprisingly, we find that the bootstrap will typically fail in this context. More specifically, we study residual-based bootstrap procedures for the LIML estimator, which attains consistency under much weaker conditions than TSLS in the current asymptotic framework. We first consider a standard residual-based bootstrap, in which the residuals of the structural-form equation are obtained by using LIML, and in which the residuals of the reduced-form equation are obtained by using the least squares estimator. We analytically demonstrate that this procedure is not asymptotically valid in that it cannot consistently estimate the limit distribution of LIML. In particular, when l is of the same order of magnitude as r_n, the bootstrap analogue of LIML converges at the same rate as the original LIML, but the bootstrap limit distribution has an asymptotic variance different from the original one. The inconsistency becomes even more severe when $l/r_n \to \infty$; in this case, the bootstrap analogue converges at a rate faster than the original LIML. The foremost reason of this bootstrap failure is that the standard procedure cannot mimic well the identification strength in the original sample. We also consider the restricted efficient (RE) bootstrap procedure of Davidson and MacKinnon (2008, 2010, 2014) that generates bootstrap data under the null (Restricted) and uses an efficient estimator of the coefficient of the reduced-form equation (Efficient). With a relatively small number of instruments, Davidson and MacKinnon (2008, 2010, 2014) show that this bootstrap procedure performs very well relative to the standard bootstrap procedure. Here, we establish that in the current context the RE bootstrap is also invalid in general. However, we also find that it effectively mimics more key parameters in the limit distribution of LIML than the standard bootstrap, and hence exhibits relatively less distortion in finite samples. Finally, we propose a modified RE bootstrap procedure and justify that it provides valid distributional approximation for LIML under many/many weak instruments. More precisely, we modify the RE bootstrap by accurately re-scaling the residuals and by introducing an alternative reduced-form estimator to help the bootstrap to mimic well the identification strength in the sample. In the simulations, the modified RE procedure is the bootstrap method that performs best overall, essentially removing the finite sample distortions generated by the standard/RE bootstraps; it also greatly outperforms the CSE-based asymptotic normal approximation.

To the best of our knowledge, this paper is the first to theoretically study the bootstrap validity in the context of IV regression under many/many weak instrument asymptotic framework. Using this alternative asymptotic framework, we obtain interesting implications of the properties of bootstrap methods that can be overlooked under conventional asymptotics. In particular, our findings highlight a fragility of bootstrap-based distributional approximations to IV estimators with respect to the number and quality of available instruments. Indeed, conditions much more restrictive than those for the normal approximation are necessary for existing bootstrap methods to work under many/many weak instruments. Our results also include a new, valid bootstrap-based inference procedure for IV models, which is able to effectively mimic the important features in the limiting distribution of interest and hence exhibits superior finite sample behavior.

The remainder of the paper is organized as follows. Section 2 introduces the model and the assumptions, provides a summary of the asymptotic theory for the LIML estimator and the CSE. Section 3 analyzes various residual-based bootstrap procedures and documents the inconsistency of the standard and RE bootstraps under many/many weak instrument sequences. Furthermore, we show that our modified bootstrap procedure provides a valid distributional approximation for LIML in this context. Section 4 contains the Monte Carlo results, and Section 5 concludes. All proofs are relegated to the Appendix.

3.2 Model, Assumptions and Asymptotic Theory

We consider a standard linear instrumental variable regression given by

$$y = X\beta + \epsilon, \tag{3.1}$$

$$X = Z\Pi + V, \tag{3.2}$$

where y and X are, respectively, an $n \times 1$ vector and an $n \times k$ matrix of observations on the endogenous variables, and Z is an $n \times l$ matrix of observations on the instruments, which we treat as deterministic. ϵ and V are, respectively, an $n \times 1$ vector and an $n \times k$ matrix of random disturbances. Also denote $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z = I_n - P_Z$, where I_n is an identity matrix with dimension n. Throughout this paper, we consider the case where k, the dimension of β, is small relative to n, but we let $l \to \infty$ as $n \to \infty$ to model the effect of having many/many weak instruments.

The model and data are assumed to satisfy the following conditions.

Assumption 1

(i) The errors $\eta_i = (\epsilon_i, V_i')'$ are i.i.d. for $i = 1, ..., n$ with mean zero and positive definite variance matrix $\Sigma = \begin{pmatrix} \sigma_{\epsilon\epsilon} & \sigma_{V\epsilon}' \\ \sigma_{V\epsilon} & \Sigma_{VV} \end{pmatrix}$; ϵ_i and V_i have finite eighth moments.

(ii) Z includes among its columns a vector of ones, $rank(Z) = l$, $\sum_{i=1}^{n} (1 - P_{ii})^2 / n \geq C > 0$ where P_{ii} denotes the diagonal elements of the matrix P_Z.

Assumption 1 (i) includes moment existence and homoscedasticity assumptions. As pointed out by Hansen, Hausman, and Newey (2008), both consistency of the LIML estimator and the CSE with many/many weak instruments depend on the homoscedasticity assumption. The condition $\sum_{i=1}^{n} (1 - P_{ii})^2 / n \geq C$ in Assumption 1 (ii) implies that $l/n \leq 1 - C$, because $P_{ii} \leq 1$ implies $\sum_{i=1}^{n} (1 - P_{ii})^2 / n \leq \sum_{i=1}^{n} (1 - P_{ii}) / n = 1 - l/n$.

Assumption 2

As $n \to \infty$, $\lambda_n = l/n \to \lambda$ for some constant λ satisfying $0 \leq \lambda < 1$. There exists a non-decreasing sequence of positive real numbers r_n such that, as $n \to \infty$, $r_n \to \infty$ and $r_n/n \to \kappa$ for some constant κ, with $0 \leq \kappa < \infty$, and such that $\Pi' Z' Z \Pi / r_n \to \Psi$, where Ψ is a positive definite matrix. Also assume that $\sqrt{l}/r_n \to 0$ and $\sum_{i=1}^{n} \|\Pi' Z_i\|^4 / r_n^2 \to 0$ as $n \to \infty$.

Assumption 2 adopts the many/many weak instruments asymptotic framework in which the number of instruments is allowed to go to infinity with the sample size. Note that r_n can be interpreted as the rate at which the concentration parameter, $\Sigma_{VV}^{-1/2}\Pi' Z' Z\Pi\Sigma_{VV}^{-1/2}$, grows as n increases. Given that the concentration parameter is a natural measure of instrument strength, one can characterize the quality of instruments by the order of magnitude of r_n, so that the slower is the divergence of r_n, the weaker are the instruments. If $r_n = n$, then the number of instruments, l, may grow as fast as n and still satisfy Assumption 2. This case corresponds to the many instrument sequence considered by Kunitomo (1980), Morimune (1983), Bekker (1994), Donald and Newey (2001), Anderson, Kunitomo and Matsushita (2010), Kuersteiner and Okui (2010) among others. Allowing l to grow and r_n to grow more slowly than n corresponds to the many weak instrument sequence considered by Chao and Swanson (2005), Stock and Yogo (2005), Hansen, Hausman, and Newey (2008), Chao, Swanson, Hausman, Newey, and Woutersen (2012), Hausman, Newey, Woutersen, Chao, and Swanson (2012), etc.

We emphasize that the many weak instrument asymptotics considered here is very different from the so-called weak instrument asymptotics, in which the number of instruments is assumed to be fixed and in which the instruments are weak in the Staiger and Stock (1997) sense (i.e., Π is specified to be in an $n^{-1/2}$ shrinking neighborhood of zero). It is well known that under this weak instrument asymptotics, the k-class IV estimators, including LIML, are inconsistent and that Wald type inferences based on these estimators can have serious size distortions. In contrast, as has been demonstrated by Chao and Swanson (2005), under many weak instrument sequence, the consistency of these estimators is attained as long as the concentration parameter increases fast enough relative to the number of instruments. For example, consider a special case where there is only one endogenous regressor, where the instruments are orthonormal ($Z'Z = nI_l$), and where we have the local-to-zero parametrization, $\Pi = n^{-\zeta}\iota_l$, $\iota_l = (1, \ldots, 1)$. In this case, $\Pi' Z' Z\Pi = n^{1-2\zeta} l \equiv r_n$. Therefore, even when the instruments are weak in the Staiger and Stock (1997) sense ($\zeta = 1/2$), the consistency of LIML requires only that the number of instruments grows to infinity[1]. This illustrates the potential benefit of using many instruments in situations where each individual instrument is only weakly correlated with the endogenous regressor. Even if each component of Π is small, the combined effect of using a large number of instruments may nevertheless allow the concentration parameter to grow sufficiently fast, so that consistent estimation can be achieved as $l, n \to \infty$.

Now we turn to describe our estimator of interest and the CSE. The k-class formulation of the LIML estimator reads

$$\hat{\beta} = \left(X' P_Z X - \hat{\alpha} X' X\right)^{-1} \left(X' P_Z y - \hat{\alpha} X' y\right),$$

with $\hat{\alpha} = \min_{\|\alpha\|=1} \frac{\alpha' Y' P_Z Y \alpha}{\alpha' Y' Y \alpha}$ and $Y = [y, X]$. We focus on LIML because it is more robust to the number and the quality of the instruments than is TSLS, the other commonly used IV estimator. It is well known in the literature that TSLS is seriously biased when the number of instruments is large. Moreover, in the current context of many weak instruments, LIML is consistent as long as r_n grows faster than $\sqrt{l}$ while TSLS is consistent only when r_n grows faster than l. In addition, LIML enjoys some asymptotic optimal properties under many instrument sequence, as has been shown by Anderson, Kunitomo, and Matsushita (2010).

Following Hansen, Hausman, and Newey (2008), Chao et al. (2012) and Hausman et al. (2012), we also distinguish between two cases depending on the speed of growth of l relative to r_n:

$$\begin{aligned} Case\,(I) &\;:\; l/r_n \to \gamma, 0 \le \gamma < \infty, \\ Case\,(II) &\;:\; l/r_n \to \infty. \end{aligned}$$

This is necessary because the convergence rates and the limiting distributions

[1]In current example, the condition $\sqrt{l}/r_n \to 0$ as $n \to \infty$ is equivalent to $n^{1-2\zeta}\sqrt{l} \to \infty$. When $\zeta = 1/2$, it is satisfied as long as $l \to \infty$.

of LIML differ in these two cases. In fact, Hansen, Hausman, and Newey (2008) have shown the following asymptotic distributional results:

Theorem 3.2.1. *Suppose Assumptions 1-2 hold. Then, in case (I)*

$$\sqrt{r_n}(\hat{\beta}-\beta) \to^d N(0,\Lambda_I), \tag{3.3}$$

where $\Lambda_I = H^{-1}\Upsilon_I H^{-1}$, $H = (1-\lambda)\Psi$, $\Upsilon_I = (1-\lambda)\sigma_{\epsilon\epsilon}\{H+\gamma\Sigma_{\tilde{V}\tilde{V}}\} + (1-\lambda)\sqrt{\gamma}\{A+A'\} + \gamma B$,

$A = E\left(\epsilon_i^2\tilde{V}_i\right) \times lim_{n\to\infty}\sum_{i=1}^n \frac{\Pi' Z_i(P_{ii}-\lambda_n)}{\sqrt{lr_n}}$;

$B = (\phi-\lambda)E\left(\left(\epsilon_i^2-\sigma_{\epsilon\epsilon}\right)\tilde{V}_i\tilde{V}_i'\right)$;

$\Sigma_{\tilde{V}\tilde{V}} = E\left(\tilde{V}_i\tilde{V}_i'\right)$, $\tilde{V} = V - \epsilon q'$, $q = \sigma_{V\epsilon}/\sigma_{\epsilon\epsilon}$ *and* $\phi = lim_{n\to\infty}\sum_{i=1}^n P_{ii}^2/l$; *in case (II),*

$$\frac{r_n}{\sqrt{l}}(\hat{\beta}-\beta) \to^d N(0,\Lambda_{II}), \tag{3.4}$$

where $\Lambda_{II} = H^{-1}\Upsilon_{II}H^{-1}$, $\Upsilon_{II} = (1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}} + B$.

Notice that in the formula of the asymptotic variance, H corresponds to the variance term that appears in conventional asymptotics, and the term with $\Sigma_{\tilde{V}\tilde{V}}$ corresponds to the additional term due to the effect of having many/many weak instruments. Thus, Case (I) with $\gamma > 0$ can be considered to be a knife-edge case where the additional variance term is of the same order as the usual variance term. If r_n grow faster than l ($\gamma = 0$), the usual variance term will dominate. On the other hand, in Case (II) r_n grows at a rate slower than l, then the additional variance term will dominate. The terms A and B account for the non-normality adjustment terms. As has been noted by Hansen, Hausman, and Newey (2008), these terms will tend to be quite small in practice.

Based on the results in Theorem 3.2.1, Hansen, Hausman, and Newey (2008) give the CSE, which are an extension of Bekker (1994)'s standard errors to the case of non-Gaussian disturbances. Let $\epsilon(\beta) = y - X\beta$, $\hat{\sigma}_{\epsilon\epsilon}(\beta) = \epsilon(\beta)'\epsilon(\beta)/(n-k)$, $\hat{\lambda}(\beta) = \epsilon(\beta)'P_Z\epsilon(\beta)/\epsilon(\beta)'\epsilon(\beta)$, $\hat{X} = P_Z X$, $\tilde{X}(\beta) = X -$

$\epsilon(\beta)\left(\epsilon(\beta)'X\right)/\epsilon(\beta)'\epsilon(\beta)$, $\hat{V}(\beta) = M_Z\tilde{X}(\beta)$, $\lambda_n = l/n$, $\phi_n = \sum_{i=1}^n P_{ii}^2/l$,

$$\begin{aligned}
\hat{H}(\beta) &= X'P_ZX - \hat{\lambda}(\beta)X'X, \\
\hat{\Upsilon}_{bkk}(\beta) &= \hat{\sigma}_{\epsilon\epsilon}(\beta)\left\{(1-\hat{\lambda}(\beta))^2\tilde{X}(\beta)'P_Z\tilde{X}(\beta) + \hat{\lambda}(\beta)^2\tilde{X}(\beta)'M_Z\tilde{X}(\beta)\right\}, \\
\hat{\Upsilon}(\beta) &= \hat{\Upsilon}_{bkk}(\beta) + \hat{A}(\beta) + \hat{A}(\beta)' + \hat{B}(\beta), \\
\hat{A}(\beta) &= \sum_{i=1}^n \left(P_{ii} - \lambda_n\right)\hat{X}_i\left(\sum_{j=1}^n \epsilon_j(\beta)^2\hat{V}_j(\beta)/n\right)', \\
\hat{B}(\beta) &= \frac{l\left(\phi_n - \lambda_n\right)}{n(1-2\lambda_n+\lambda_n\phi_n)}\sum_{i=1}^n\left(\epsilon_i(\beta)^2 - \hat{\sigma}_{\epsilon\epsilon}(\beta)\right)\hat{V}_i(\beta)\hat{V}_i(\beta)'.
\end{aligned}$$

Their asymptotic variance estimator is given by

$$\hat{\Lambda} = \hat{H}^{-1}\hat{\Upsilon}\hat{H}^{-1},\ \hat{H} = \hat{H}(\hat{\beta}),\ \hat{\Upsilon} = \hat{\Upsilon}(\hat{\beta}).$$

Notice that $\hat{H}^{-1}\hat{\Upsilon}_{bkk}(\hat{\beta})\hat{H}^{-1}$ is identical to the Bekker (1994) variance estimator. The other terms in $\hat{\Upsilon}(\hat{\beta})$ account for third and fourth moment terms that are present with some forms of nonnormality. In the case of many weak instruments, $\hat{\Lambda}$ can be quite different from the usual asymptotic variance estimator $\hat{\sigma}_{\epsilon\epsilon}\hat{H}^{-1}$ because when the reduced form R^2 is small, $\hat{\Upsilon}$ can become much larger than $\hat{H}$. Then, the asymptotic normality of the t-test based on the CSE can be established

$$t_{cse} = \frac{c'(\hat{\beta}-\beta)}{\sqrt{c'\hat{\Lambda}c}} \to^d N(0,1) \tag{3.5}$$

where $c'\hat{\beta}$ is a linear combination of LIML and c is the linear combination coefficient. Although in practice one cannot distinguish between Cases (I) and (II) since r_n is unobservable, this would not be a problem because Wald inferences can be implemented using the t-test statistic in eq.(3.5) irrespective of Cases (I) or (II).

However, the CSE-based normal approximation can be inaccurate in samples of moderate size due to the slower than $n^{-1/2}$ convergence speed of $\hat{\beta}$ in the context of many weak instruments. In fact, as can be seen from simulation evidence in Hansen, Hausman, and Newey (2008) and also in ours, Wald-type inferences based on the asymptotic normal approximation can have serious size distortions, especially when the degree of endogeneity is high (i.e., when the correlation between ϵ_i and V_i is high). This provides a motivation for the use of the bootstrap as an alternative method of inference. In particular, we can improve the quality of inference by relying on the bootstrap instead of a normal approximation when computing critical values for test statistics. Moreover, the CSE have a rather tedious form that empirical researchers might find difficult to implement. This also motivates the use of bootstrap-based methods. In cases where the analytical standard errors have a tedious form or are believed to be difficult to estimate, the bootstrap often provides a useful empirical alternative. For example, the widely used percentile-type $100(1-\alpha)\%$ symmetric confidence interval (CI)s take the form

$$CI = \left[c'\hat{\beta} - q^*_{1-\alpha},\, c'\hat{\beta} + q^*_{1-\alpha}\right],$$

where $q^*_{1-\alpha}$ is such that $P^*\left(\left|c'\left(\hat{\beta}^* - \hat{\beta}\right)\right| \leq q^*_{1-\alpha}\right) = 1-\alpha$ and P^* denotes the probability measure induced by bootstrap. One can thus implement inference using percentile-type CIs when they are valid without actually computing the analytic standard errors. Another approach also many times employed in the literature is to directly estimate the variance-covariance matrix of $\hat{\beta}$ using the bootstrap, as an alternative to an analytic standard-errors estimator. More specifically, this approach leads to the following $100(1-\alpha)\%$ CIs:

$$CI = \left[c'\hat{\beta} - z_{1-\alpha}\sqrt{c'\widehat{\Lambda}^*_{boot}c},\, c'\hat{\beta} + z_{1-\alpha}\sqrt{c'\widehat{\Lambda}^*_{boot}c}\right],$$

where $z_{1-\alpha}$ is such that $P\left(|Z| \leq z_{1-\alpha}\right) = 1 - \alpha$ with $Z \sim N\left(0,1\right)$, and

$$\widehat{\Lambda}^*_{boot} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\beta}^*_b - \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^*_b\right)\left(\hat{\beta}^*_b - \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}^*_b\right)'$$

where B is chosen large enough so that $\widehat{\Lambda}^*_{boot}$ approximates well the variance-covariance matrix of interest. However, as we shall see in the following section, these two approaches will typically become invalid under many/many weak instruments if bootstrap data are generated by existing procedures.

3.3 Main Results

In this section, we study the asymptotic validity of the bootstrap methods when applied to the LIML estimator. Three residual-based bootstrap methods adapted to the linear IV model are considered. We begin with what we call the standard bootstrap procedure that amounts to re-sampling the residuals obtained by using the LIML estimate for equation (1) and the least squares estimate for the reduced-form equation (2) to generate bootstrap data. Then, we consider the restricted efficient (RE) bootstrap procedure of Davidson and MacKinnon (2008, 2010, 2014) which generates bootstrap data under the null hypothesis $H_0 : \beta = \beta_0$ and uses an efficient estimator of the coefficient of the reduced-form equation. We demonstrate that these two bootstrap procedures fail to provide valid distributional approximation to LIML under many/many weak instruments. Furthermore, we propose a modified version of the RE bootstrap procedure, and we prove the bootstrap consistency of this procedure.

The following notations are used for the bootstrap asymptotics (see Chang and Park (2003b) for similar notation and for several useful bootstrap asymptotic properties): for any bootstrap statistic T^* we write $T^* \to^{P^*} 0$ in probability if for any $\delta > 0$, $\epsilon > 0$, $\lim_{n\to\infty} P\left[P^*\left(|T^*| > \delta\right) > \epsilon\right] = 0$, i.e.,

$P^*(|T^*| > \delta) = o_P(1)$. Also, we write $T^* = O_{P^*}(n^\varphi)$ in probability if and only if for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $\lim_{n\to\infty} P[P^*(|n^{-\varphi}T^*| > M_\delta) > \delta] = 0$, i.e., for any $\delta > 0$ there exists a $M_\delta < \infty$ such that $P^*(|n^{-\varphi}T^*| > M_\delta) = o_P(1)$. Finally, we write $T^* \to^{d^*} T$ in probability if, conditional on the sample, T^* weakly converges to T under P^*, for all samples contained in a set with probability converging to one. Specifically, we write $T^* \to^{d^*} T$ in probability if and only if $E^*(f(T^*)) \to E(f(T))$ in probability for any bounded and uniformly continuous function f. To be concise, we sometimes use the short version $T^* \to^{P^*} 0$ to say that $T^* \to^{P^*} 0$ in probability, and use $T^* = O_{P^*}(n^\varphi)$ for $T^* = O_{P^*}(n^\varphi)$ in probability.

3.3.1 Standard bootstrap procedure

We begin with the standard residual-based bootstrap procedure. Given the LIML estimate of β and the least squares (first-stage) estimate $\hat{\Pi} = (Z'Z)^{-1}Z'X$, the residuals are obtained as

$$\hat{\epsilon} = y - X\hat{\beta} \tag{3.6}$$

$$\hat{V} = X - Z\hat{\Pi} \tag{3.7}$$

Then, the residual $\hat{\epsilon}$ is re-centered to yield $\tilde{\epsilon}$ and (ϵ^*, V^*) are drawn from the empirical distribution function of $\left(\tilde{\epsilon}, \hat{V}\right)$. Notice that we do not re-center $\hat{V}$ here since it already has mean zero by our assumption that Z includes a constant term. Next, we set

$$y^* = X^*\hat{\beta} + \epsilon^* \tag{3.8}$$

$$X^* = Z\hat{\Pi} + V^* \tag{3.9}$$

Finally, we compute the bootstrap analogue of LIML using the pseudo-sample $\{X^*, y^*\}$

$$\hat{\beta}^*_{std} = \left(X^{*\prime} P_Z X^* - \hat{\alpha}^* X^{*\prime} X^*\right)^{-1} \left(X^{*\prime} P_Z y^* - \hat{\alpha}^* X^{*\prime} y^*\right),$$

where $\hat{\alpha}^* = \min_{\|\alpha\|=1} \frac{\alpha' Y^{*\prime} P_Z Y^* \alpha}{\alpha' Y^{*\prime} Y^* \alpha}$ and $Y^* = [y^*, X^*]$. [2]

Below we show that the standard residual-based bootstrap fails to consistently estimate the limiting distribution of LIML in both Cases (I) and (II).

Theorem 3.3.1. *Suppose Assumptions 1-2 hold. Then, in case (I) with* $0 < \gamma < \infty$*,*

$$\sqrt{r_n}(\hat{\beta}^*_{std} - \hat{\beta}) \to_{d^*} N(0, \bar{\Lambda}_I) \text{ in probability},$$

where

$$\begin{aligned}
\bar{\Lambda}_I &= \bar{H}_I^{-1} \bar{\Upsilon}_I \bar{H}_I^{-1} \\
\bar{\Upsilon}_I &= (1-\lambda)\sigma_{\epsilon\epsilon} \left\{\bar{H}_I + \gamma \bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\bar{A} + \bar{A}'\right\} + \gamma \bar{B} \\
\bar{H}_I &= H + (1-\lambda)\gamma \Sigma_{VV} \\
\bar{\Sigma}_{\tilde{V}\tilde{V}} &= \Sigma_{\tilde{V}\tilde{V}} - \lambda\left\{\Sigma_{VV} + (\lambda - 2)\sigma_{V\epsilon}\sigma'_{V\epsilon}/\sigma_{\epsilon\epsilon}\right\} \\
\bar{A} &= (1-\lambda)A \\
\bar{B} &= (1 - 2\lambda + \lambda\phi)B + \lambda(\phi - \lambda)^2 \left\{2E\left(\epsilon_i^3 \tilde{V}_i\right) q' + q\left[E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right] q'\right\};
\end{aligned}$$

in case (II),

$$\sqrt{l}(\hat{\beta}^*_{std} - \hat{\beta}) \to^{d^*} N(0, \bar{\Lambda}_{II}) \text{ in probability},$$

[2] This procedure is called Unrestricted Inefficient (UI) procedure in Davidson and MacKinnon (2008) since the bootstrap d.g.p. is not generated under the null hypothesis (thus Unrestricted) and the least squares estimator $\hat{\Pi}$, instead of a more efficient estimator, is used to estimate the reduced-form coefficient (thus Inefficient).

where

$$\begin{aligned}\bar{\Lambda}_{II} &= \bar{H}_{II}^{-1}\bar{\Upsilon}_{II}\bar{H}_{II}^{-1} \\ \bar{\Upsilon}_{II} &= (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_{II}+\bar{\Sigma}_{\tilde{V}\tilde{V}}\right\}+\bar{B} \\ \bar{H}_{II} &= (1-\lambda)\Sigma_{VV}\end{aligned}$$

Theorem 3.3.1 makes it easy to quantify the inconsistency of the bootstrap approximation in the case of many/many weak instruments. According to Theorem 3.3.1, in Case (I) the bootstrap analogue $\hat{\beta}^*_{std}$ has the same convergence rate as $\hat{\beta}$ and the bootstrap distribution is asymptotically normal, but its asymptotic variance-covariance matrix is different from the one derived in Theorem 3.2.1. More specifically, the formula of the asymptotic variance-covariance matrix of LIML consists of certain key parameters such as H, which characterizes the identification strength in the IV model, and also various moments of the disturbances $(\epsilon_i, V_i')'$. It turns out that the standard bootstrap fails to mimic well these key parameters, and thus cannot provide a valid approximation to the limiting distribution of LIML. To see why the bootstrap fails in the current context, let us first consider the term H. The LIML objective function with bootstrap pseudo-data reads $\widehat{Q}^*(\beta) = (y^* - X^*\beta)' P_Z (y^* - X^*\beta) / (y^* - X^*\beta)' (y^* - X^*\beta)$, and the usual Taylor expansion of the first-order condition $\partial\widehat{Q}^*(\hat{\beta}^*_{std})/\partial\beta = 0$ yields

$$\hat{\beta}^*_{std} - \hat{\beta} = \left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)^{-1} \partial\widehat{Q}^*(\hat{\beta})/\beta$$

where $\bar{\beta}^*$ is an intermediate value on the line joining $\hat{\beta}^*_{std}$ and $\hat{\beta}$. We can show that in Case (I), the bootstrap Hessian term

$$\left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)/r_n \rightarrow^{P^*} \bar{H}_I = (1-\lambda)\left(\Psi + \gamma\Sigma_{VV}\right) = H + (1-\lambda)\gamma\Sigma_{VV}$$

in probability; P^* denotes the probability measure induced by the standard bootstrap. Thus, in contrast to the limit of the original Hessian term in Theorem 3.2.1, the bootstrap approximation results in an extra term $(1-\lambda)\gamma\Sigma_{VV}$. Similarly, by applying a bootstrap CLT, we obtain that although the bootstrap Jacobian term $\sqrt{r_n}\left(\partial\widehat{Q}^*(\hat{\beta})/\partial\beta\right)$ converges in probability to normal distribution, the term H in the asymptotic variance of $\sqrt{r_n}\left(\partial\widehat{Q}(\beta_0)/\partial\beta\right)$ is also replaced by $H+(1-\lambda)\gamma\Sigma_{VV}$. Intuitively, in the current context, the standard bootstrap fails to adequately capture the identification strength (or instrument strength) in the original sample. Indeed, "extra" identification strength will be generated by the bootstrap d.g.p. as long as the rate of growth of l is not slower than that of r_n, and this "extra" identification strength will result in the extra term $(1-\lambda)\gamma\Sigma_{VV}$.

The bootstrap failure becomes even more severe in case (II). It can be seen from Theorem 3.3.1 that in this case, the convergence rate of $\hat{\beta}^*_{std}$ turns out to be different of that of $\hat{\beta}$: $\hat{\beta}^*_{std}-\hat{\beta}=O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ in probability, while $\hat{\beta}-\beta=O_P\left(\frac{\sqrt{l}}{r_n}\right)$. Notice that $\frac{1}{\sqrt{l}}/\frac{\sqrt{l}}{r_n}\to 0$ since $l/r_n\to\infty$ in Case (II). That is, $\hat{\beta}^*_{std}$, the bootstrap analogue of LIML, will converge to $\hat{\beta}$, the true value in the bootstrap world, at a higher speed than $\hat{\beta}$ converges to β_0. In addition, the formula of the bootstrap asymptotic variance $\bar{\Lambda}_{II}$ differs greatly from Λ_{II} in Theorem 3.2.1. Indeed, the conventional variance term H does not appear in the formula of Υ_{II} because in Case (II) it is dominated by the many/many weak instrument adjustment term $(1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}}$ and by the non-normality adjustment term B. In contrast, $\bar{H}_{II}$, the bootstrap analogue of H in Case (II), does appear in the formula of $\bar{\Upsilon}_{II}$. This is also because of the "extra" identification strength generated by the bootstrap d.g.p., which guarantees that $\bar{H}_{II}$ will not be dominated by the other terms.

Furthermore, some algebra shows that by the re-sampling scheme of the standard bootstrap, the following results hold for the bootstrap disturbances

$\left(\epsilon_i^*, V_i^{*\prime}\right)'$ under the current asymptotic framework:

$$\begin{aligned} E^*\left(V_i^* \epsilon_i^*\right) &= n^{-1} \sum_{i=1}^{n} \hat{V}_i \tilde{\varepsilon}_i \quad \to^P \quad (1-\lambda)\sigma_{V\varepsilon} \\ E^*\left(V_i^* V_i^{*\prime}\right) &= n^{-1} \sum_{i=1}^{n} \hat{V}_i \hat{V}_i' \quad \to^P \quad (1-\lambda)\Sigma_{VV} \end{aligned}$$

Therefore, except for $\sigma_{\epsilon\epsilon}$, the standard residual bootstrap fails to consistently estimate the other elements in the variance-covariance matrix of $(\epsilon_i, V_i')'$ when the number of instruments grows at the same speed as the sample size ($\lambda \neq 0$). This also implies that for $\tilde{V}_i^* \equiv V_i^* - \epsilon_i^* \frac{E^*\left(V_i^* \epsilon_i^*\right)'}{E^*\left(\epsilon_i^{*2}\right)}$,

$$\begin{aligned} E^*\left(\tilde{V}_i^* \tilde{V}_i^{*\prime}\right) &= E^*\left(V_i^* - \epsilon_i^* \frac{E^*\left(V_i^* \epsilon_i^*\right)'}{E^*\left(\epsilon_i^{*2}\right)}\right)\left(V_i^* - \epsilon_i^* \frac{E^*\left(V_i^* \epsilon_i^*\right)'}{E^*\left(\epsilon_i^{*2}\right)}\right)' \\ &\to^P \Sigma_{\tilde{V}\tilde{V}} - \lambda\left(\Sigma_{VV} + (\lambda-2)\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right), \end{aligned}$$

where E^* denotes the expectation under the probability measure induced by the standard bootstrap. As long as l goes to infinity at the same rate as n, the standard residual bootstrap will fail to provide, even asymptotically, a good approximation to $\Sigma_{\tilde{V}\tilde{V}}$, which plays an important role in the formula of the asymptotic variance in Theorem 3.2.1. Similar results of inconsistency can also be shown for other bootstrap moments such as $E^*\left(\epsilon_i^{2*} \tilde{V}_i^*\right)$ and $E^*\left(\epsilon_i^{*2} \tilde{V}_i^* \tilde{V}_i^{*\prime}\right)$.

Remarks:

1. Since $\hat{\beta}$ attains consistency under our Assumptions 1-2, folklore may suggest that the bootstrap d.g.p. in (6)-(9) will be valid as long as $\hat{\Pi}$ is a consistent estimator of Π. Interestingly, this turns out to be wrong according to Theorem 3.3.1. Indeed, it is shown in Portnoy (1984) that $\hat{\Pi}$ will be consistent provided that the growth rate of l is not too fast related to the growth rate of the sample size ($l(logl)/n \to 0$). However, we can see

from Theorem 3.3.1 that without proper restriction on the relationship between l and r_n, such a condition is not adequate to guarantee the bootstrap consistency under many/many weak instrument sequences.

2. Similar results of bootstrap failure can be shown for other k-class IV estimators such as the commonly used TSLS estimator, the bias-adjusted TSLS estimator (Nagar (1959), Rothenberg (1984), etc.), and the modified LIML estimator proposed by Fuller (1977). We omit these results for the conciseness of the paper.

3. Instead of using residual-based bootstrap, one may consider implementing the nonparametric i.i.d. bootstrap (pairs bootstrap), which resamples the rows of the matrix $(y : X : Z)$. Indeed, (y_i^*, X_i^*, Z_i^*), the ith row of each bootstrap sample, is simply one of the row of $(y : X : Z)$, chosen at random with probability $1/n$. However, the extensive simulation evidence in Davidson and MacKinnon (2010,2014) shows that the pairs bootstrap performs substantially worse than residual-based bootstrap methods, even when the number of instruments is kept small relative to the sample size. Note that under the nonparametric i.i.d. bootstrap, the bootstrap analogue of the slope coefficient in the first-stage regression is characterized by

$$E^* \left(Z_i^* Z_i^{*'} \right)^{-1} E^* \left(Z_i^* X_i^* \right) = \left(n^{-1} Z'Z \right)^{-1} \left(n^{-1} Z'X \right),$$

which is exactly $\hat{\Pi}$ used in the standard residual-based bootstrap.

On the other hand, when l is of lower order of magnitude relative to r_n, the formula of the asymptotic variance of $\sqrt{r_n}(\hat{\beta} - \beta)$ is considerably simplified

$$\sqrt{r_n}(\hat{\beta} - \beta) \to_d N\left(0, \sigma_{\epsilon\epsilon} \Psi^{-1}\right)$$

and the standard residual-based bootstrap does consistently estimate this asymptotic distribution.

Corollary 3.3.1. *Suppose that Assumptions 1-2 holds and suppose that*

$l/r_n \to 0$ ($\gamma = 0$), then

$$sup_{x\in R}\left|P^*\left(\sqrt{r_n}(\hat{\beta}^*_{std}-\hat{\beta})\leq x\right)-P\left(\sqrt{r_n}(\hat{\beta}-\beta)\leq x\right)\right|\to^P 0$$

where P^* denotes the probability measure induced by the standard bootstrap procedure.

Closely related to our paper is the literature on bootstrapping linear model with increasing dimension. More precisely, consider the following model

$$y_i = X_i'\beta + \epsilon_i,\ \ i = 1, ..., n$$

where X_i's and β are p-dimensional vectors, and ϵ_i's are i.i.d. errors. Asymptotics where p may increase with n have been considered by Bickel, Freedman, Bickel, Doksum, and Hodges (1983), Portnoy (1984), Mammen (1989, 1993), among others. In particular, Bickel, Freedman, Bickel, Doksum, and Hodges (1983) show that residual-based bootstrap consistently estimate the distribution of the least square estimates if $p^2/n \to 0$, and for linear contrasts of the least square estimates it works if $p/n \to 0$. Mammen (1988) generalizes these results to the case of M estimates. Furthermore, it is shown by these authors that under large p asymptotics, residual-based bootstrap even works in the case that asymptotic normal approximation typically fails. Apparently, the rate of growth of p with respect to n is crucial for bootstrap consistency under this large p asymptotic framework.

In contrast, we show that for the current IV model with large l, the bootstrap consistency depends importantly on the relative magnitude of r_n vis-à-vis l as $n \to \infty$, but not so much on the relationship between l and n. Additionally, different from bootstrapping under large p asymptotics, conditions more restrictive than those for the normal approximation are necessary for the standard bootstrap to work under many/many weak instruments. For the orthonormal instruments example in Section 2, since $\Pi'Z'Z\Pi = n^{1-2\zeta}l = r_n$, the bootstrap consistency requires $l/r_n = n^{2\zeta-1} \to 0$

($\zeta < 1/2$, the instruments need to be stronger than the weak instruments in the Staiger and Stock (1997) sense), while the CSE-based normal approximation only requires $\sqrt{l}/r_n = l^{-1/2} n^{2\zeta-1} \to 0$.

3.3.2 Restricted Efficient Bootstrap Procedure

In this section, we study the other residual-based bootstrap procedure recently proposed by Davidson and MacKinnon (2008, 2010, 2014) for the IV model. The RE residual-based bootstrap has two key features. First, the bootstrap pseudo-data is obtained under the null $H_0 : \beta = \beta_0$ (instead of using the LIML estimate $\hat{\beta}$). Second, the RE bootstrap uses a more efficient (reduced-form) estimate instead of $\hat{\Pi}$ in the standard residual bootstrap. Following Davidson and MacKinnon (2008, 2010, 2014), we first obtain the residuals for the RE procedure by

$$\begin{aligned} \epsilon(\beta_0) &= y - X\beta_0 \\ \tilde{V}(\beta_0) &= X - Z\tilde{\Pi}(\beta_0) \end{aligned}$$

where

$$\tilde{\Pi}(\beta_0) = (Z'Z)^{-1} Z' \left(X - \epsilon(\beta_0) \frac{\epsilon'(\beta_0) M_Z X}{\epsilon'(\beta_0) M_Z \epsilon(\beta_0)} \right).$$

Then, (ϵ^*, V^*) are drawn from the empirical distribution function of

$$\left(\sqrt{\frac{n}{n-k}} \epsilon(\beta_0), \sqrt{\frac{n}{n-l}} \tilde{V}(\beta_0) \right)$$ [3].

[3] Since $\left(\sqrt{\frac{n}{n-k}} \epsilon(\beta_0), \sqrt{\frac{n}{n-l}} \tilde{V}(\beta_0) \right)$ is not necessarily mean zero, (ϵ^*, V^*) should be drawn from the empirical distribution function of $\left(\sqrt{\frac{n}{n-k}} \left(\epsilon(\beta_0) - \bar{\epsilon}(\beta_0) \right), \sqrt{\frac{n}{n-l}} \left(\tilde{V}(\beta_0) - \bar{\tilde{V}}(\beta_0) \right) \right)$ where, $\bar{\epsilon}(\beta_0) \equiv \frac{1}{n} \sum_{i=1}^{n} \epsilon_i(\beta_0)$ and $\bar{\tilde{V}}(\beta_0) \equiv \frac{1}{n} \sum_{i=1}^{n} \tilde{V}_i(\beta_0)$.

Next, we set

$$\begin{aligned} y^* &= X^*\beta_0 + \epsilon^* \\ X^* &= Z\tilde{\Pi}(\beta_0) + V^* \end{aligned}$$

and obtain $\hat{\beta}^*_{re}$ using pseudo-data generated by this procedure. Notice that $\tilde{\Pi}(\beta_0)$ is the maximum likelihood estimator of Π when β is constrained to take the null value β_0. It is also used in Kleibergen (2002) and Moreira (2003) to construct their weak identification robust statistics. In particular, they show that using $\tilde{\Pi}(\beta_0)$ rather than $\hat{\Pi}$ leads to their Lagrange Multiplier (LM) test for $H_0 : \beta = \beta_0$ that is asymptotically pivotal even under weak instruments asymptotics of Staiger and Stock (1997).

The RE bootstrap has been applied very successfully in IV models with relatively small number of instruments. As can be observed from the extensive simulation results in Davidson and MacKinnon (2008, 2010), using this procedure instead of the standard residual bootstrap or the nonparametric i.i.d. bootstrap greatly improves size control for testing the null hypothesis $H_0 : \beta = \beta_0$, especially when the instruments are relatively weak (e.g., when $a = 2$ in Davidson and MacKinnon (2008, 2010), which corresponds to the case where the concentration parameter equals 4). The RE bootstrap is also used in Davidson and MacKinnon (2014) to build confidence sets for β in a similar context. Their simulation results show that in contrast to what is widely believed, even when the instruments are quite weak, it is possible to make the Wald-based confidence sets perform well using the RE bootstrap procedure.

However, we find that under many/many weak instrument sequences the RE bootstrap is also invalid in general. The following theorem states the asymptotic distributional results for the RE bootstrap.

Theorem 3.3.2. *Suppose Assumptions 1-2 hold. Then, in Case (I) with*

$0 < \gamma < \infty$ and under $H_0 : \beta = \beta_0$,

$$\sqrt{r_n}(\hat{\beta}^*_{re} - \beta_0) \to_{d^*} N(0, \tilde{\Lambda}_I) \text{ in probability,}$$

where

$$\begin{aligned}
\tilde{\Lambda}_I &= \tilde{H}_I^{-1}\tilde{\Upsilon}_I\tilde{H}^{-1} \\
\tilde{\Upsilon}_I &= (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_I + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\tilde{A} + \tilde{A}'\right\} + \gamma\tilde{B} \\
\tilde{H}_I &= H + (1-\lambda)\gamma\Sigma_{\tilde{V}\tilde{V}} \\
\tilde{A} &= \sqrt{1-\lambda}A \\
\tilde{B} &= \frac{1-2\lambda+\lambda\phi}{1-\lambda}B
\end{aligned}$$

In case (II) and under $H_0 : \beta = \beta_0$,

$$\sqrt{l}(\hat{\beta}^*_{re} - \beta_0) \to^{d^*} N(0, \tilde{\Lambda}_{II}) \text{ in probability,}$$

where

$$\begin{aligned}
\tilde{\Lambda}_{II} &= \tilde{H}_{II}^{-1}\tilde{\Upsilon}_{II}\tilde{H}_{II}^{-1} \\
\tilde{H}_{II} &= (1-\lambda)\Sigma_{\tilde{V}\tilde{V}} \\
\tilde{\Upsilon}_{II} &= (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_{II} + \Sigma_{\tilde{V}\tilde{V}}\right\} + \tilde{B}
\end{aligned}$$

Investigating the results in Theorem 3.3.2, we find that the RE bootstrap is also inconsistent as long as l goes to infinity at a rate equal to or faster than that of r_n. For example, using similar arguments as for the standard residual bootstrap, we obtain that in Case (I) and under H_0, the RE-based approximation of the Hessian term

$$\left(\partial^2\widehat{Q}^*(\bar{\beta}^*)/\partial\beta\partial\beta'\right)/r_n \to^{P^*} \tilde{H}_I = H + (1-\lambda)\gamma\Sigma_{\tilde{V}\tilde{V}} \tag{3.10}$$

in probability; P^* denotes the probability measure induced by the RE bootstrap and $\bar{\beta}^*$ denotes an intermediate value on the line joining $\hat{\beta}^*_{re}$ and β_0. Thus, the RE bootstrap also fails to adequately mimic the instrument strength in the sample and results in an approximation error of the same order of magnitude as the key parameter H. A similar problem occurs in the bootstrap Jacobian term $\sqrt{r_n}\left(\partial\widehat{Q}^*(\beta_0)/\partial\beta\right)$. In Case (II), the bootstrap failure becomes more severe as $\hat{\beta}^*_{re}$ also converges at a rate of $O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, the same convergence rate as $\hat{\beta}^*_{std}$. For the moments of bootstrap disturbances, some algebra shows that under $H_0 : \beta = \beta_0$,

$$\begin{aligned} E^*\left(V_i^*\epsilon_i^*\right) &= \sqrt{\frac{n}{n-k}}\sqrt{\frac{n}{n-l}}\left(\frac{1}{n}\sum_{i=1}^n \tilde{V}_i(\beta_0)\epsilon_i(\beta_0)\right) \to^P \frac{\sigma_{V\epsilon}}{\sqrt{1-\lambda}} \\ E^*\left(V_i^*V_i^{*'}\right) &= \frac{n}{n-l}\left(\frac{1}{n}\sum_{i=1}^n \tilde{V}_i(\beta_0)\tilde{V}_i'(\beta_0)\right) \to^P \Sigma_{VV} + \frac{\lambda}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}} \end{aligned}$$

where E^* denotes the expectation under the probability induced by the RE bootstrap. Similar bootstrap inconsistency also appears in the non-normality adjustment terms A and B.

Interestingly, the RE bootstrap does consistently estimate $\Sigma_{\tilde{V}\tilde{V}}$, the variance of residuals from the population regression of V_i on ϵ_i; that is, we can show that by the RE bootstrap d.g.p.,

$$E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) = E^*\left(V_i^*V_i^{*'}\right) - \left\{E^*\left(V_i^*\epsilon_i^*\right)E^*\left(V_i^*\epsilon_i^*\right)'/E^*\left(\epsilon_i^{*2}\right)\right\} \to^P \Sigma_{\tilde{V}\tilde{V}},$$

under H_0, including the case where l is of the same order of magnitude as n. This is remarkable since according to the formula of the asymptotic variance in Theorem 3.2.1, the many/many weak instruments adjustment term crucially depends on $\Sigma_{\tilde{V}\tilde{V}}$. As has been highlighted by Hansen, Hausman, and Newey (2008), in practice this adjustment term can be comparable to the usual asymptotic variance term H, while the non-normality adjustment terms will tend to be very small relative to $\Sigma_{\tilde{V}\tilde{V}}$ and H. Furthermore, although the RE bootstrap cannot also consistently estimate H, it holds that

$H \leq \tilde{H}_I \leq \bar{H}_I$ in Case (I) and $\tilde{H}_{II} \leq \bar{H}_{II}$ in Case (II) since $\Sigma_{\tilde{V}\tilde{V}} \leq \Sigma_{VV}$ by the definition of $\tilde{V}$. Therefore, our asymptotic results in Theorem 3.3.2 predict that the RE-based distributional approximation for LIML will typically be more precise than the standard bootstrap-based approximation and this is indeed confirmed by our simulation.

It is also easy to show that, as the standard residual bootstrap, the RE bootstrap is consistent when $l/r_n \to 0$.

Corollary 3.3.2. *Suppose that Assumptions 1-2 holds and that $l/r_n \to 0$ ($\gamma = 0$), then under $H_0 : \beta = \beta_0$,*

$$sup_{x \in R} \left| P^* \left(\sqrt{r_n}(\hat{\beta}^*_{re} - \beta_0) \leq x \right) - P \left(\sqrt{r_n}(\hat{\beta} - \beta_0) \leq x \right) \right| \to^P 0$$

where P^* denotes the probability measure induced by the RE bootstrap procedure.

Thus, similar to the standard bootstrap, the RE bootstrap is asymptotically valid only when the available instruments are sufficiently strong so that the concentration parameter grows at a faster rate than the number of instruments. One can thus expect the performance of the RE-based distributional approximation for LIML to be quite sensitive to the quality and number of instruments. In the next section, we propose a modified RE procedure which is able to consistently estimate the distribution of LIML under much weaker conditions on the growth rate of l relative to r_n.

3.3.3 Modified RE Bootstrap Procedure

In this section, we propose a modified version of the RE bootstrap d.g.p. so that the approximation errors can be removed from the bootstrap limit distribution. The modified RE (MRE) procedure achieves this goal by correctly re-scaling the residuals and by using an alternative reduced-form estimator so that the bootstrap can effectively mimic the identification strength in the original sample.

More specifically, for the MRE bootstrap the residuals are obtained as

$$\begin{aligned} \epsilon(\beta_0) &= y - X\beta_0 \\ \hat{V} &= X - Z\hat{\Pi}, \end{aligned}$$

Then, (ϵ_m^*, V_m^*) is drawn from the empirical distribution function of

$$\left(\sqrt{\frac{n}{n-l}} M_Z \epsilon(\beta_0), \sqrt{\frac{n}{n-l}} \hat{V}\right).$$

To generate the bootstrap d.g.p., we use $\tilde{\Pi}_m(\beta_0)$ which is based on modifying the RE reduced-form estimator $\tilde{\Pi}(\beta_0)$ in the following way

$$\tilde{\Pi}_m(\beta_0) = \tilde{\Pi}(\beta_0)\left(\tilde{\Psi}^{-1/2}(\beta_0)\tilde{\Psi}_m^{1/2}(\beta_0)\right) \tag{3.11}$$

where

$$\begin{aligned} \tilde{\Psi}(\beta_0) &= \tilde{\Pi}'(\beta_0) Z' Z \tilde{\Pi}(\beta_0) \\ \tilde{\Psi}_m(\beta_0) &= \left(\tilde{\Psi}(\beta_0) - l\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0), 0\right)^+ \\ \widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0) &= \frac{1}{n-l}\tilde{X}'(\beta_0) M_Z \tilde{X}(\beta_0) \\ \tilde{X}(\beta_0) &= X - \epsilon(\beta_0)\left(\frac{\epsilon' M_Z(\beta_0) X}{\epsilon'(\beta_0) M_Z \epsilon(\beta_0)}\right) \end{aligned}$$

and $(\cdot, 0)^+ = max\,(\cdot, 0)$. Next, we set

$$\begin{aligned} y^* &= X^*\beta_0 + \epsilon_m^* \\ X^* &= Z\tilde{\Pi}_m(\beta_0) + V_m^* \end{aligned}$$

and compute $\hat{\beta}_m^*$ using the pseudo-data obtained by this bootstrap procedure.

Notice that an alternative modified procedure based on $\hat{\beta}$ could also be pursued. More precisely, this procedure amounts to using $\hat{\epsilon}$ instead of $\epsilon(\beta_0)$ and using $\tilde{\Pi}_m(\hat{\beta})$ instead of $\tilde{\Pi}_m(\beta_0)$ when generating the bootstrap d.g.p. Under $H_0 : \beta = \beta_0$, these two procedures are asymptotically equivalent. However, trial simulation shows that the modified bootstrap procedure generated under β_0 typically has much better finite sample performance, especially when the instruments are weak.

To motivate the MRE procedure, let us first consider the modification introduced in eq.(3.11). Since the bootstrap generated under $\tilde{\Pi}(\beta_0)$ cannot mimic well the instrument strength in the current context and results in an approximation error of order at least as large as the concentration parameter, we introduce $\tilde{\Psi}^{-1/2}(\beta_0)\tilde{\Psi}_m^{1/2}(\beta_0)$ as a correction factor to remove the "extra" instrument strength in the RE bootstrap d.g.p. Indeed, we can show that under the null $H_0 : \beta = \beta_0$,

$$\frac{\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)}{r_n} = \frac{\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)}{r_n} - \left(\frac{l}{r_n}\right)\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0) \to^P \Psi$$

in both Cases (I) and (II). Therefore, in contrast to the standard/RE bootstrap, the modified procedure is able to consistently estimate the key parameter Ψ that characterizes the identification strength under many/many weak instruments.

Also, consider the bootstrap disturbances generated by the MRE procedure. With our approach of re-scaling the residuals into $\left(\sqrt{\frac{n}{n-l}}M_Z\epsilon(\beta_0), \sqrt{\frac{n}{n-l}}\hat{V}\right)$, the MRE bootstrap is able to mimic well each component of the covariance matrix of $(\epsilon_i, V_i')'$ even in the case that l is of the same order of magnitude as n:

$$
\begin{aligned}
E^*\left(\epsilon_{i,m}^{*2}\right) &= \frac{n}{n-l}\left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{\epsilon\epsilon} \\
E^*\left(V_{i,m}^{*}\epsilon_{i,m}^{*}\right) &= \frac{n}{n-l}\left(\frac{\hat{V}'M_Z\epsilon(\beta_0)}{n}\right) \to^P \sigma_{V\epsilon} \\
E^*\left(V_{i,m}^{*}V_{i,m}^{*\prime}\right) &= \frac{n}{n-l}\left(\frac{\hat{V}'\hat{V}}{n}\right) \to^P \Sigma_{VV}
\end{aligned}
$$

under $H_0 : \beta = \beta_0$, where E^* denote the expectation generated by the MRE procedure.

Remark: Our correction factor in eq.(3.11) is related to the restricted-efficient-corrected (REC) bootstrap in Davidson and MacKinnon (2008, pg.458). The REC bootstrap is motivated by the fact that $\tilde{a}^2 = \tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)/\left(n^{-1}\tilde{V}'(\beta_0)\tilde{V}(\beta_0)\right)$, their RE-based estimator of the concentration parameter (in $k = 1$ case), is inconsistent under Staiger and Stock (1997)'s weak instrument asymptotics, and has a bias of $l \times (1 - \sigma_{V\epsilon}^2/\sigma_{\epsilon\epsilon})$. Although consistent estimation of the concentration parameter is impossible under weak instrument asymptotics, an unbiased estimator can be constructed as $\tilde{a}_{BC}^2 = (0, \tilde{a}^2 - l(1-\tilde{\rho}^2))^+$ where $\tilde{\rho}^2 = \epsilon'(\beta_0)\tilde{V}(\beta_0)/\left\{(\epsilon'(\beta_0)\epsilon(\beta_0))\left(\tilde{V}'(\beta_0)\tilde{V}(\beta_0)\right)\right\}^{1/2}$, and the reduced-form equation of the REC bootstrap d.g.p. can be generated using

$$X^* = Z\tilde{\Pi}(\beta_0)\left(\frac{\tilde{a}_{BC}}{\tilde{a}}\right) + V^*$$

Under current many/many weak instrument sequences, it can be shown that in Case (I)

$$\begin{aligned}\frac{\tilde{\Pi}'_{BC}(\beta_0)Z'Z\tilde{\Pi}_{BC}(\beta_0)}{r_n} &= \frac{\tilde{\Pi}'(\beta_0)Z'Z\tilde{\Pi}(\beta_0)}{r_n} \\ &\quad -\frac{l}{r_n}\left\{\frac{\tilde{V}'(\beta_0)\tilde{V}(\beta_0)}{n} - \left(\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n}\right)^{-1}\left(\frac{\epsilon'(\beta_0)\tilde{V}(\beta_0)}{n}\right)^2\right\} \\ &\to^P \Psi + \gamma\sigma_{\tilde{V}\tilde{V}} - (1-\lambda)\gamma\sigma_{\tilde{V}\tilde{V}} = \Psi + \lambda\gamma\sigma_{\tilde{V}\tilde{V}}\end{aligned}$$

Thus, the REC bootstrap will be inconsistent when $\lambda \neq 0$. In Case (II), due to this inconsistency $\tilde{\Pi}'_{BC}(\beta_0)Z'Z\tilde{\Pi}_{BC}(\beta_0)/r_n$ will diverge to infinity, leading the REC-based bootstrap analogue of LIML to also converge too fast, like the standard/RE bootstrap analogues.

Below, we introduce some additional assumptions that help to simplify the variance formula, and we show that our modified bootstrap procedure is able to consistently estimate the limiting distribution of LIML under either of these assumptions.

Assumption 3

(i) $\lambda_n \to \lambda \neq 0$ and $n^{-1}\sum_{i=1}^n |P_{ii} - \lambda_n| \to 0$ as $l, n \to \infty$.

(ii) $\lambda_n \to \lambda = 0$ as $l, n \to \infty$.

Assumption 3 (i) is also used in Anatolyev and Gospodinov (2011) for many instrument sequence and in Anatolyev (2012) for many regressor sequence. As has been pointed out in their papers, this condition allows that the number of instruments increases at the same rate as the sample size but requires that (almost) all diagonal elements of the projection matrix P_Z converge to λ (note that under conventional asymptotics they converge to zero), and it will typically hold if the instruments are homogenous across i. On the other hand, Assumption 3 (ii) requires that the number of instruments grows at a slower rate than the sample size. This case is most important in empirical applications, especially in microeconomic studies, where the number of instruments is usually small relative to the sample size. For example, in their study of return to schooling problem, Donald and Newey (2001) and Hansen, Hausman, and Newey (2008) used 180 instruments with a sample size of $329,509$. The variance formula of LIML will be simplified under Assumption 3 (i) or Assumption 3 (ii) because the non-normality adjustment terms A and B in the variance formula disappear under either assumption, that is, $A = B = 0$. Such assumptions are mild because among the various terms in the asymptotic variance of LIML, A and B will typically be quite small compared with other terms, as emphasized by Hansen, Hausman, and Newey (2008).

The distributional results for the MRE bootstrap procedure are stated in the following theorem.

Theorem 3.3.3. *Suppose that Assumptions 1-2 hold. Also suppose either Assumption 3(i) or Assumption 3(ii) holds. Then under* $H_0 : \beta = \beta_0$, *in Case (I),*

$$\sup_{x \in R} \left| P^* \left(\sqrt{r_n}(\hat{\beta}_m^* - \beta_0) \leq x \right) - P \left(\sqrt{r_n}(\hat{\beta} - \beta_0) \leq x \right) \right| \to^P 0$$

and in Case (II),

$$\sup_{x \in R} \left| P^* \left(\frac{r_n}{\sqrt{l}}(\hat{\beta}_m^* - \beta_0) \leq x \right) - P \left(\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta_0) \leq x \right) \right| \to^P 0$$

where P^* *denotes the probability measure induced by the MRE bootstrap procedures.*

Theorem 3.3.3 states that the MRE bootstrap procedure mimics well the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta)$ in case (I) and the limiting distribution of $\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta)$ in case (II), thus giving the asymptotic validity of percentile type CIs constructed based on the MRE bootstrap. In terms of finite sample behavior, our simulation evidence in Section 4 shows that the MRE-based percentile type CIs typically have better empirical coverage rates compared with normal approximation-based CIs.

3.3.4 Bootstrapping t-test with Corrected Standard Error

In view of the success of the MRE procedure in providing distributional approximation for LIML, we can expect that the distribution of t_{cse}, t-test statistic based on the LIML estimate and the CSE, can also be well approximated by our bootstrap procedure. Moreover, since t_{cse} is asymptotically standard normal under many/many weak instrument asymptotics, folklore suggests that the standard and RE bootstraps should also be capable of consistently estimating its distribution even if these bootstrap procedures cannot adequately mimic the limit distribution of LIML. This conjecture turns out to be correct, because one can show that in Case (I),

$$r_n\widehat{\Lambda}^*_{std} \approx V^*\left[\sqrt{r_n}\left(\hat{\beta}^*_{std}-\hat{\beta}\right)\right] \text{ and } r_n\widehat{\Lambda}^*_{re} \approx V^*\left[\sqrt{r_n}\left(\hat{\beta}^*_{re}-\beta_0\right)\right]$$

with "$A \approx B$" being shorthand for $A^{-1}B \to^{P^*} I_k$ in probability and V^* denoting the variance computed under the corresponding bootstrap distribution. $\widehat{\Lambda}^*_{std}$ and $\widehat{\Lambda}^*_{re}$ denote the CSEs computed using the pseudo-data generated by the standard and RE bootstrap d.g.p., respectively. Then, the result of weak convergence in probability for the bootstrap analogues of t_{cse} can be established, i.e.,

$$\frac{c'(\hat{\beta}^*_{std}-\hat{\beta})}{\sqrt{c'\widehat{\Lambda}^*_{std}c}} = \frac{c'\sqrt{r_n}(\hat{\beta}^*_{std}-\hat{\beta})}{\sqrt{c'r_n\widehat{\Lambda}^*_{std}c}} \to^{d^*} N(0,1)$$

in probability and similar result holds for the RE bootstrap. Analogously, we have for Case (II)

$$l\widehat{\Lambda}^*_{std} \approx V^* \left[\sqrt{l}\left(\hat{\beta}^*_{std} - \hat{\beta}\right)\right] \text{ and } l\widehat{\Lambda}^*_{re} \approx V^* \left[\sqrt{l}\left(\hat{\beta}^*_{re} - \beta_0\right)\right].$$

Thus, one can show that the standard/RE bootstrap-based approximations of the distribution of t_{cse} converges to standard normal distribution in probability, regardless of the fact that in Case (II) even the convergence speed of $\hat{\beta}^*_{std}$ and $\hat{\beta}^*_{re}$ differs from that of $\hat{\beta}$. More precisely, an application of the continuous mapping theorem for weak convergence in probability yields the following result for all the three bootstrap procedures.

Theorem 3.3.4. *Suppose that Assumptions 1-2 hold, also suppose that* $H_0 : \beta = \beta_0$ *holds for the RE and MRE bootstraps, then*

$$sup_{x \in R} \left| P^* \left(t^*_{cse} \leq x\right) - P\left(t_{cse} \leq x\right)\right| \to^P 0$$

where t^*_{cse} *denotes the t-CSE test statistic generated by one of the three bootstrap procedures.*

Theorem 3.3.4 gives asymptotic validity for percentile-t type CIs based on the standard, RE and MRE bootstrap procedures. Monte Carlo simulations show that for percentile-t CIs, all the three bootstrap procedures have reasonable empirical coverage rates with the CIs based on the MRE procedure performing best. This is not surprising considering that the bootstrap is able to provide asymptotic refinement only when the test statistic is asymptotically pivotal and when the bootstrap d.g.p. consistently estimates the original d.g.p. (see Beran (1988)). Among the three bootstrap procedures, one can only expect the MRE bootstrap to provide asymptotic refinement for percentile-t type CIs since the other two procedures are not able to consistently estimate the original d.g.p., as has been shown previously. We leave a formal study of the MRE bootstrap's higher order properties for future work.

3.4 Simulation Results

The goal of this section is to evaluate the finite sample performance of the bootstrap methods studied in the previous sections. Following Davidson and MacKinnon (2008, 2010 and 2014), we use the following DGP:

$$
\begin{aligned}
y &= \beta X + \varepsilon \\
X &= aw + v,
\end{aligned}
$$

where $w \in \mathcal{S}(Z)$, the subspace spanned by the columns of the instruments Z. w is an n-vector with $\|w\|^2 = 1$. As pointed out in their papers, the only property of Z that matters here is the subspace spanned by the columns of Z and in their setting, all the explanatory power comes from the vector w and the other columns of Z are simply noise. For the disturbances, we set

$$
\begin{aligned}
\epsilon &= r\epsilon_1 + \rho\epsilon_2 \\
v &= \epsilon_2,
\end{aligned}
$$

with $(\epsilon_1, \epsilon_2)' \sim N(0, \mathbf{I})$, $r^2 + \rho^2 = 1$. The strength of the instruments is measured by the parameter a, the square of which equals the concentration parameter.

In Figures 1-4, we present non-rejection frequencies for asymptotic and bootstrap tests. For ease of comparison, we report results of the MRE bootstrap versus the REC bootstrap in figures 5-8, separately from other bootstrap procedures. The simulation evidence are based on 1,000 replications and $B = 399$ bootstrap samples. The sample size is 100 and we use the LIML estimator throughout the simulation.

For all bootstrap procedures, we consider both percentile and percentile-t type Wald tests. Notice that we present the properties of percentile type bootstrap tests for two reasons: (i) to show that they confirm our theory which predicts that except for the MRE bootstrap, the other procedures are invalid and, (ii) to emphasize that in cases where the analytical standard errors have a tedious form or, are believe to be difficult to estimate as it is the case here, a MRE percentile bootstrap-based inference could provide a useful alternative. Let us start with the standard residual bootstrap. Percentile type bootstrap Wald tests reject $H_0 : \beta = \beta_0$ if $|\hat{\beta} - \beta_0|$ is among the $0.05(B+1)$ biggest values in

$$\left\{|\hat{\beta} - \beta_0|, |\hat{\beta}_1^* - \hat{\beta}|, ..., |\hat{\beta}_B^* - \hat{\beta}|\right\}. \tag{3.12}$$

Notice that normally, we should reject $H_0 : \beta = \beta_0$ if $\sqrt{r_n}|\hat{\beta} - \beta_0|$ is among the $0.05(B + 1)$ biggest values in

$$\left\{\sqrt{r_n}|\hat{\beta} - \beta_0|, \sqrt{r_n}|\hat{\beta}_1^* - \hat{\beta}|, ..., \sqrt{r_n}|\hat{\beta}_B^* - \hat{\beta}|\right\}$$

in Case (I) and reject $H_0 : \beta = \beta_0$ if $\frac{r_n}{\sqrt{l}}|\hat{\beta} - \beta_0|$ is among the $0.05(B + 1)$ biggest values in

$$\left\{\frac{r_n}{\sqrt{l}}|\hat{\beta} - \beta_0|, \frac{r_n}{\sqrt{l}}|\hat{\beta}_1^* - \hat{\beta}|, ..., \frac{r_n}{\sqrt{l}}|\hat{\beta}_B^* - \hat{\beta}|\right\}$$

in Case (II). However, although we do not know the exact value of r_n in practice, we are still able to use the procedure described by (3.12) since $\sqrt{r_n}$ and $r_n/\sqrt{l}$ will be canceled out in Case (I) and (II), respectively.

For percentile-t type bootstrap Wald tests, we reject $H_0 : \beta = \beta_0$ if $\frac{|\hat{\beta} - \beta_0|}{\sqrt{\hat{\Lambda}}}$ is among the $0.05(B + 1)$ biggest values in

$$\left\{\frac{|\hat{\beta} - \beta_0|}{\sqrt{\widehat{\Lambda}}}, \frac{|\hat{\beta}_1^* - \hat{\beta}|}{\sqrt{\widehat{\Lambda}_1^*}}, ..., \frac{|\hat{\beta}_B^* - \hat{\beta}|}{\sqrt{\widehat{\Lambda}_B^*}}\right\}.$$

In this formula, $\sqrt{\widehat{\Lambda}}$ is the CSE defined in Section 3.2 and $\sqrt{\widehat{\Lambda}_{std}^*}$ is its standard bootstrap counterpart.

Also, since bootstrap data of the RE, REC and MRE procedures are generated under the null, percentile type bootstrap Wald tests reject $H_0 : \beta = \beta_0$ if $|\hat{\beta} - \beta_0|$ is among the $0.05(B+1)$ biggest values in

$$\left\{ |\hat{\beta} - \beta_0|, |\hat{\beta}^*_{j,1} - \beta_0|, ..., |\hat{\beta}^*_{j,B} - \beta_0| \right\},$$

where $j \in \{re, rec, m\}$. Percentile-t type bootstrap Wald tests reject $H_0 : \beta = \beta_0$ if $\frac{|\hat{\beta}-\beta_0|}{\sqrt{\widehat{\Lambda}}}$ is among the $0.05(B+1)$ biggest values in

$$\left\{ \frac{|\hat{\beta} - \beta_0|}{\sqrt{\widehat{\Lambda}}}, \frac{|\hat{\beta}^*_{j,1} - \beta_0|}{\sqrt{\hat{\Lambda}^*_{j,1}}}, ..., \frac{|\hat{\beta}^*_{j,B} - \beta_0|}{\sqrt{\hat{\Lambda}^*_{j,B}}} \right\}$$

where $j \in \{re, rec, m\}$. Finally, asymptotic theory-based Wald (t) tests rely on critical values of the $N(0,1)$ distribution.

The first two figures each contains six plots and pertain to percentile and percentile-t type bootstrap tests. They show the effect of varying the number of instruments for three values of a^2 and two values of ρ. Specifically, we vary a^2 across rows ($a^2 \in \{4, 8, 16\}$) and ρ across columns ($\rho \in \{0.1, 0.8\}$). One can interpret $a^2 = 4$ as a very weak instruments case, $a^2 = 8$ as a weak instruments case and $a^2 = 16$ as a moderately strong instruments case. When $\rho = 0.1$, there is not much correlation between the structural and reduced-form disturbances; when $\rho = 0.8$, there is a great deal of correlation.

As in the simulation results reported in Davidson and MacKinnon (2008)[4], CSE-based asymptotic Wald (t) tests for the LIML estimator underreject when ρ is small and overreject when ρ is large, especially in the case that a^2 is small. In particular, asymptotic Wald (t) tests have noticeable finite sample distortions for $\rho = 0.8$ and $a^2 = 4$. Indeed, under this setting, the actual non-rejection rates of nominal 95% asymptotic Wald (t) tests vary between 85% and 80% for values of l between 20 and 50. These non-rejection rates

[4] See, e.g., Figure 2 and 4 in Davidson and Mackinnon (2008)

increase to around 90% to 85% when $\rho = 0.8$ and $a^2 = 8$.

Figure 1 shows clearly that percentile bootstrap tests based on the standard residual and the RE bootstrap overreject with many/many weak instruments. Also, it turns out that the distortions of both standard and RE bootstrap tests increase when the strength of the instruments decreases and/or the number of instruments increases. Thus, our results in Theorems 3.1 and 3.2 give an excellent approximation to the finite sample behavior of these bootstrap procedures. Furthermore, we find that the RE bootstrap tends to dominate the standard bootstrap, confirming our theoretical predictions in Section 3.2 that the RE bootstrap-based approximation of the distribution of LIML is typically more precise than the standard bootstrap-based approximation. On the other hand, MRE percentile bootstrap tests have much better performance for all values of a^2 and ρ. It is remarkable that in all plots, the MRE bootstrap displays very small distortions irrespective of the values of l. This is not surprising considering that the MRE bootstrap is the only bootstrap procedure able to mimic well the distribution of LIML under many/many weak instrument sequence. Also, MRE percentile bootstrap tests have large improvement over CSE-based asymptotic Wald (t) tests when $\rho = 0.8$.

Figure 2 shows that non-rejection frequencies of the standard/RE bootstrap percentile-t tests are much better than their corresponding percentile versions. These results are in line with our Theorem 3.3.4 which predicts in particular that percentile-t approximations based on these two bootstrap procedures are asymptotically valid even if their percentile counterparts are not. Also, we find that standard bootstrap tests have almost the same performance as the CSE-based asymptotic Wald (t) tests for all configuration of a^2, ρ and l. The RE bootstrap improves upon the asymptotic theory and the standard bootstrap, especially when $\rho = 0.8$, but is still notably distorted for small values of a^2. The MRE bootstrap has the best performance among all the procedures.

Figures 3-4 each contains nine plots and pertain to percentile and percentile-t type bootstrap tests. They show the effect of varying the value of ρ for three values of a^2 and three values of l. In particular, Figure 3 shows that when ρ is small, percentile type bootstrap tests based on the standard and the RE bootstraps have poor non-rejection frequencies in comparison to CSE-based asymptotic Wald (t) tests. For large values of ρ, the distortion of CSE-based asymptotic Wald (t) tests become severe while the non-rejection frequencies of RE percentile bootstrap tests become better and even improve upon asymptotic Wald (t) tests in some cases. This is natural considering that our theoretical analysis in Section 3.2 (see e.q.(3.10)) shows that the RE bootstrap approximation error depends crucially on $\Sigma_{\tilde{V}\tilde{V}}$, which equals $1 - \rho^2$ in current simulation setting. It also turns out that MRE percentile bootstrap tests perform much better than standard/RE percentile bootstrap tests, and improve upon asymptotic Wald (t) tests in most cases.

As in Figure 2, Figure 4 shows that non-rejection frequencies of percentile-t type bootstrap tests are higher than those of percentile type bootstrap tests for all the three bootstrap procedures. However, in contrast to Figure 3 where the most severe distortions occur when ρ is relatively small, the standard/RE percentile-t type bootstrap tests tend to overreject when the value of ρ becomes large, as also noticed in Figure 2 when $\rho = 0.8$. In particular, standard bootstrap tests, as well as asymptotic Wald (t), can have non-rejection frequencies as low as 75%. In contrast, non-rejection frequencies of MRE percentile-t tests are very close to 95% across all the settings of a^2, ρ and l.

We compare non-rejection rates of MRE and REC bootstrap tests in figures 5-8. Notice that the grids on the vertical lines are changed for ease of comparison. Investigating the results for percentile type bootstrap tests in Figure 5 and Figure 7, we find that REC percentile bootstrap tests typically have large distortion when l is large and ρ is small. This is also in line with our analysis in Section 3.3 which states that the distortion of the REC approximation depends mainly on the values of λ and $\Sigma_{\tilde{V}\tilde{V}}$. On the other hand, figure 6 and figure 8 shows that MRE and REC percentile-t type bootstrap tests have non-rejection rates very close to each other. Also, MRE/REC percentile-t type bootstrap tests improve substantially upon asymptotic Wald (t) tests, especially when a^2 is small and ρ is large.

3.5 Conclusion

The main contribution of this paper is to study the validity of the bootstrap for inference in linear IV regression when the available instruments may be weak and the number of instruments may be large. Using the asymptotic framework of many/many weak instruments, we obtain new theoretical results about the finite-sample behavior of the bootstrap methods that can be overlooked under the conventional asymptotic framework.

In particular, we show that a standard i.i.d. residual-based bootstrap method is unable to consistently estimate the limiting distribution of LIML under many/many weak instrument sequences. More specifically, the standard bootstrap cannot mimic well the parameter that characterizes the identification strength in the original sample. It also fails to adequately mimic certain important properties of the disturbances in the IV model. These failures lead the bootstrap distribution to converge to a limit different from the original one. Moreover, we show that the RE bootstrap proposed by Davidson and MacKinnon (2008, 2010, 2014) is also invalid in general. However, the RE bootstrap is able to effectively mimic more features in the limiting distribution of LIML, and thus its finite-sample distortion is typically smaller than that of the standard bootstrap. Finally, we propose a modified RE bootstrap and we show that this procedure provides a valid distributional approximation to LIML under many/many weak instruments. A Monte Carlo experiment shows that our procedure has outstanding small sample performance compared with asymptotic normal approximation based on the CSE and the other bootstrap procedures.

An extension of this work will include a study in general nonlinear framework on the bootstrap validity (e.g., Hall and Horowitz (1996)'s nonparametric i.i.d. bootstrap, Brown and Newey (2002)'s efficient bootstrap, etc.) for GMM and GEL estimators under many weak moment sequence proposed by Newey and Windmeijer (2009).

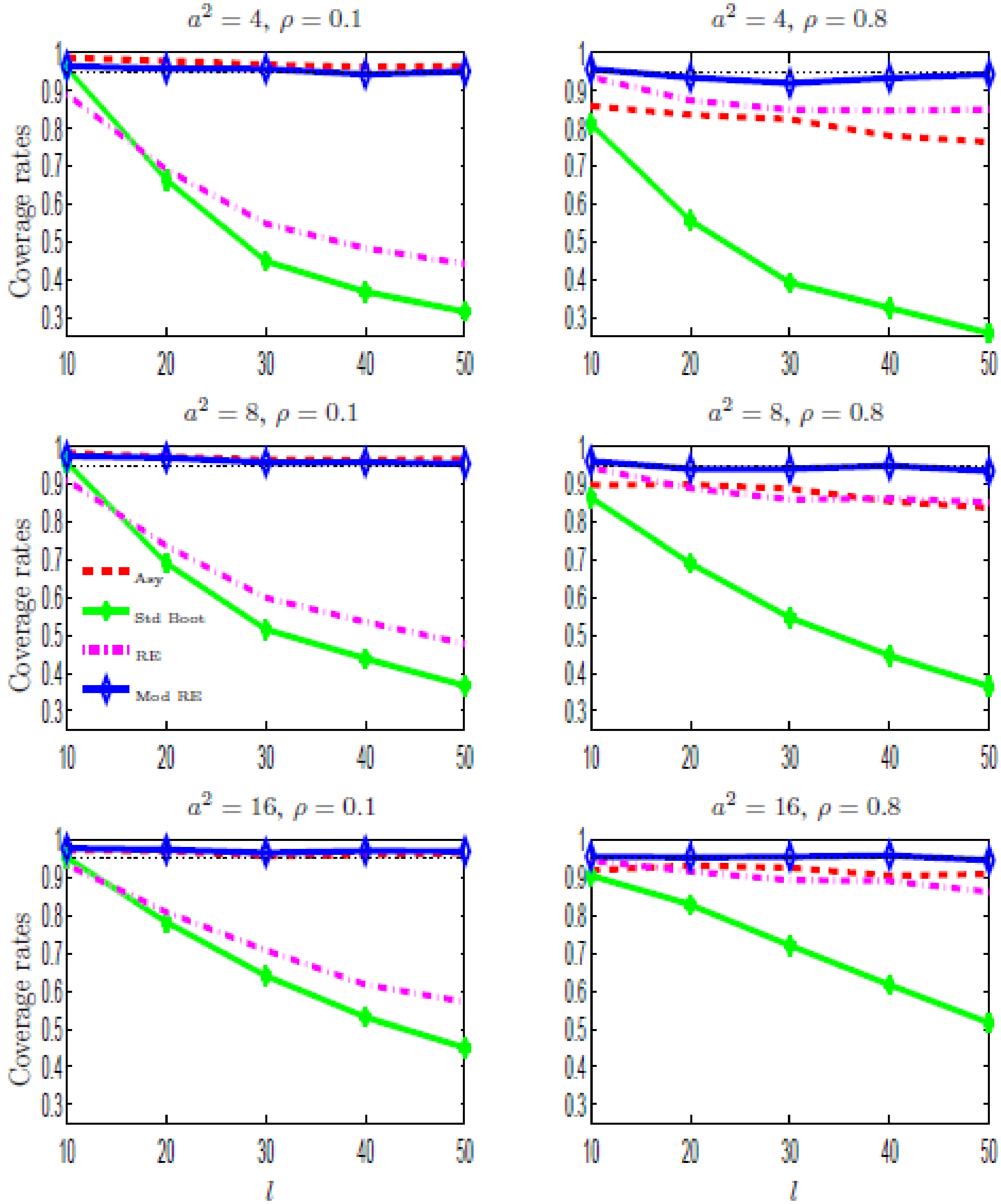

Figure 3.1: Non-rejection rates for percentile type bootstrap Wald tests as a function of l

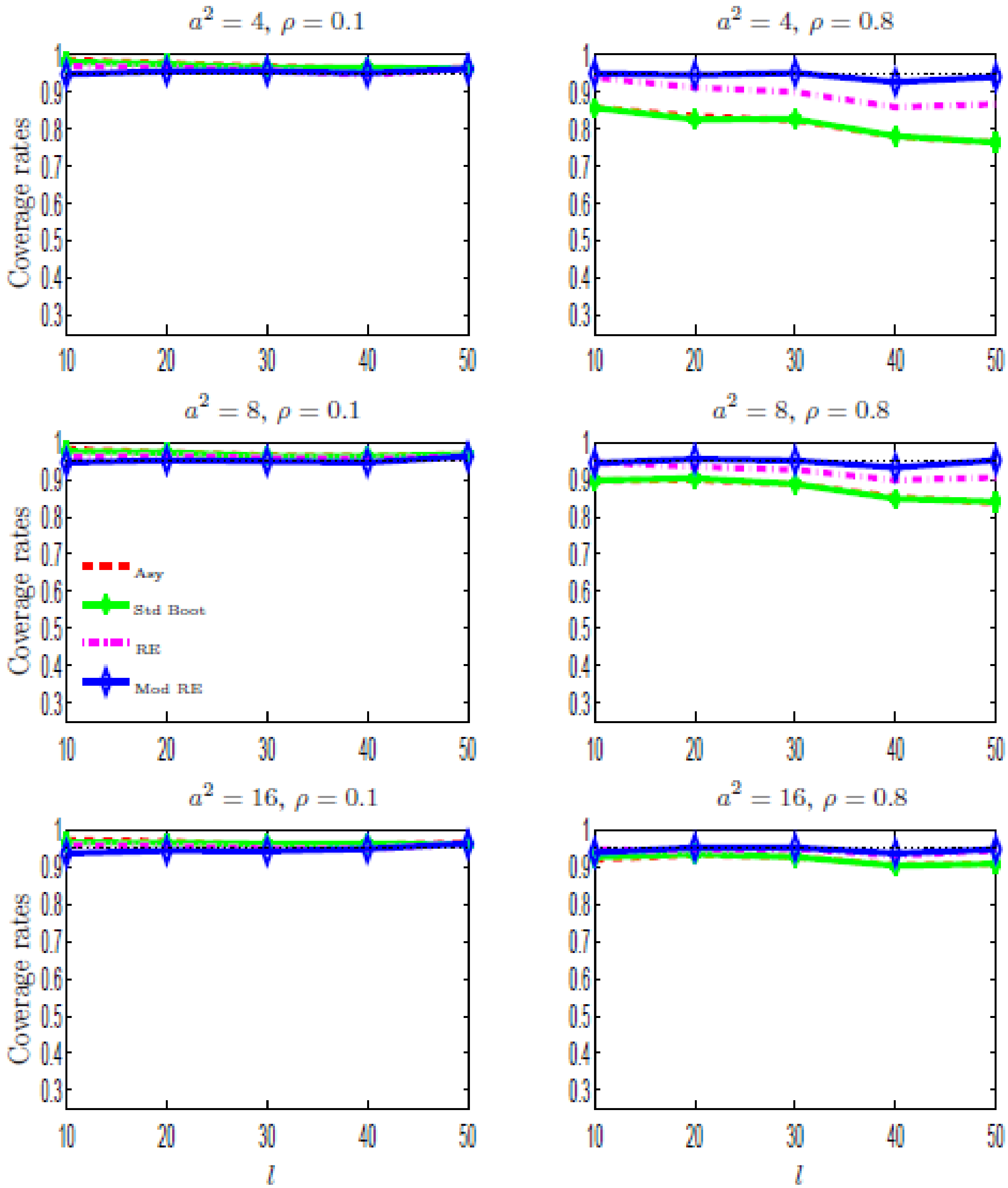

Figure 3.2: Non-rejection rates for percentile-t type bootstrap Wald tests as a function of l

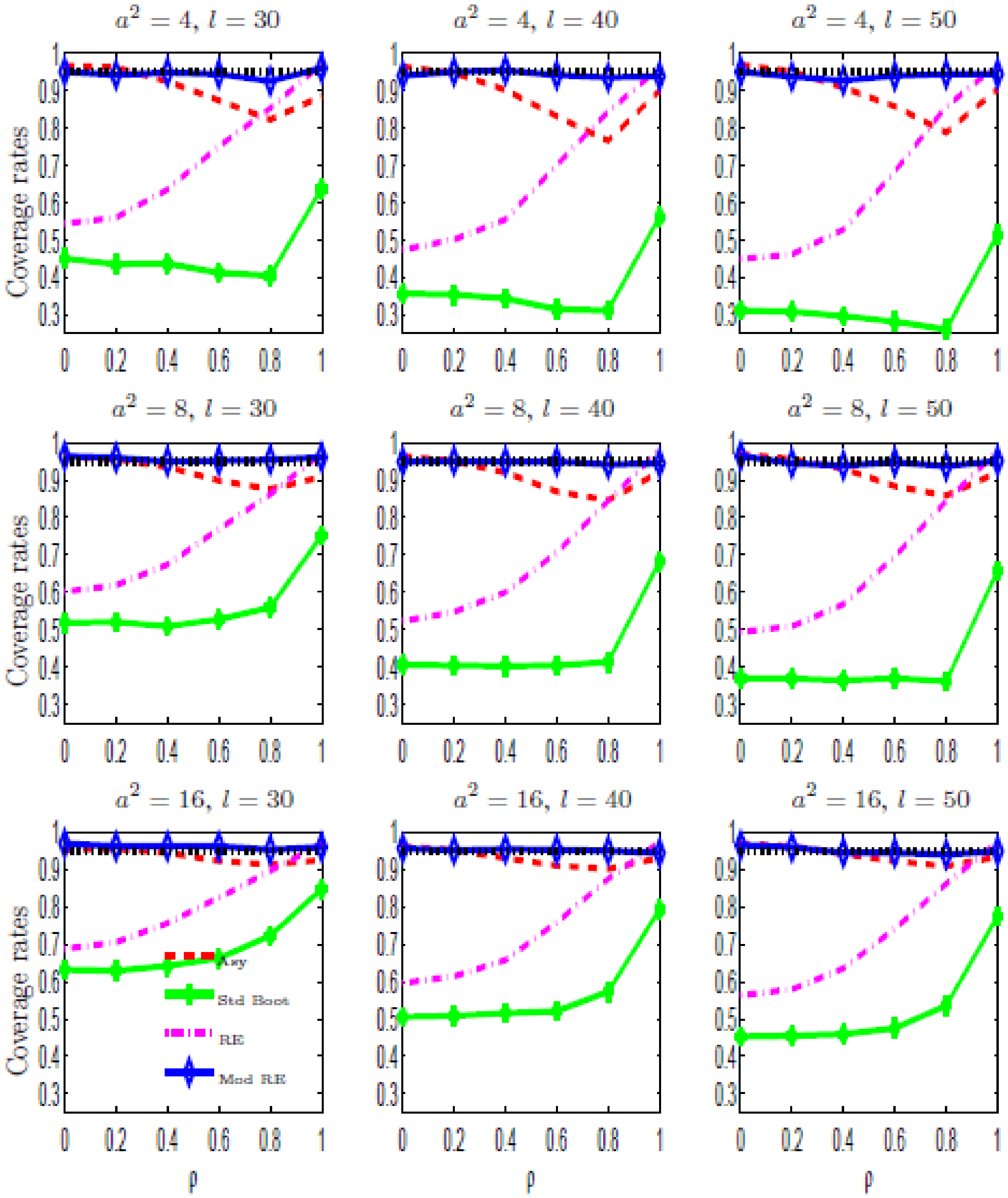

Figure 3.3: Non-rejection rates for percentile type bootstrap Wald tests as a function of ρ

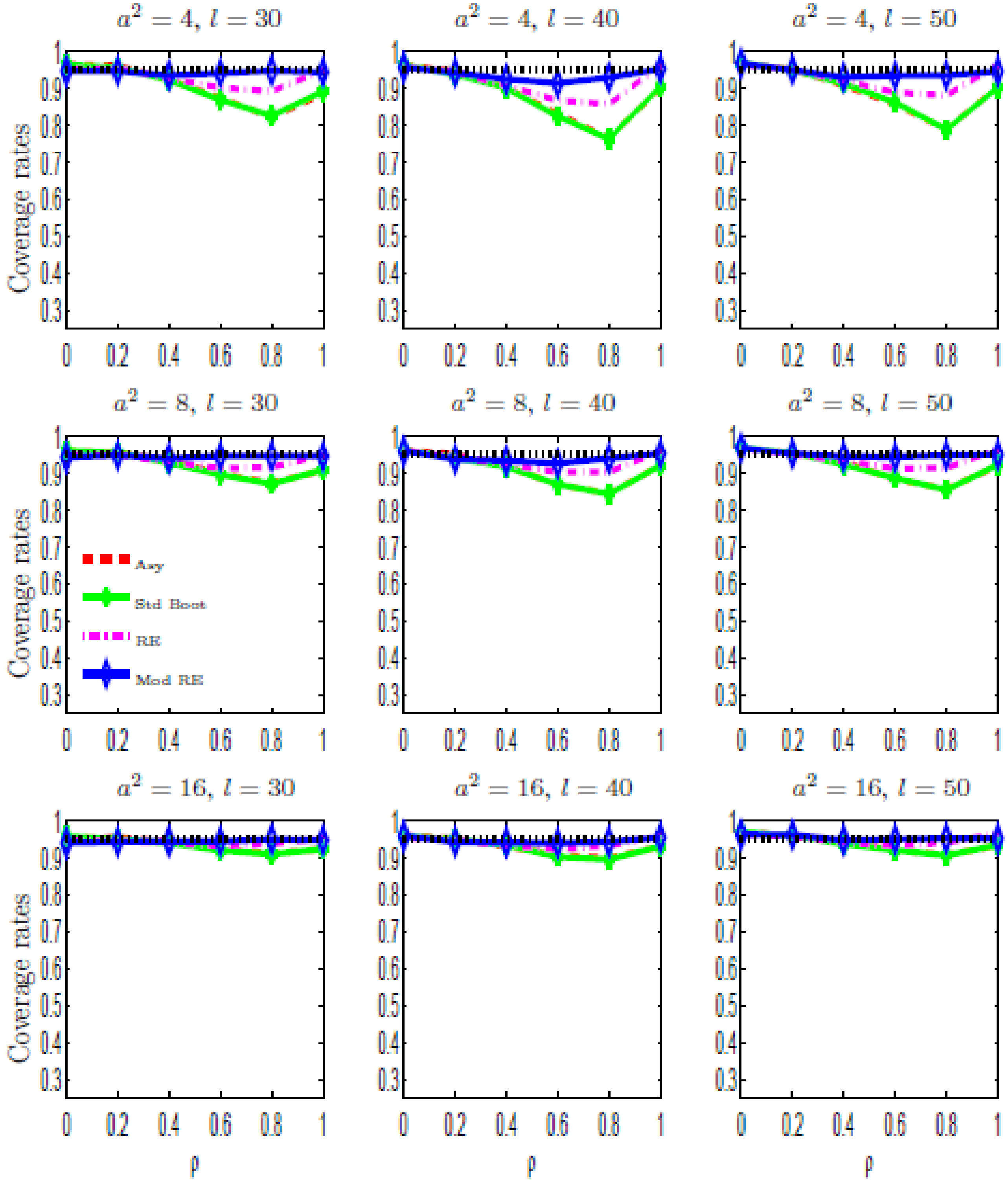

Figure 3.4: Non-rejection rates for percentile-t type bootstrap Wald tests as a function of ρ

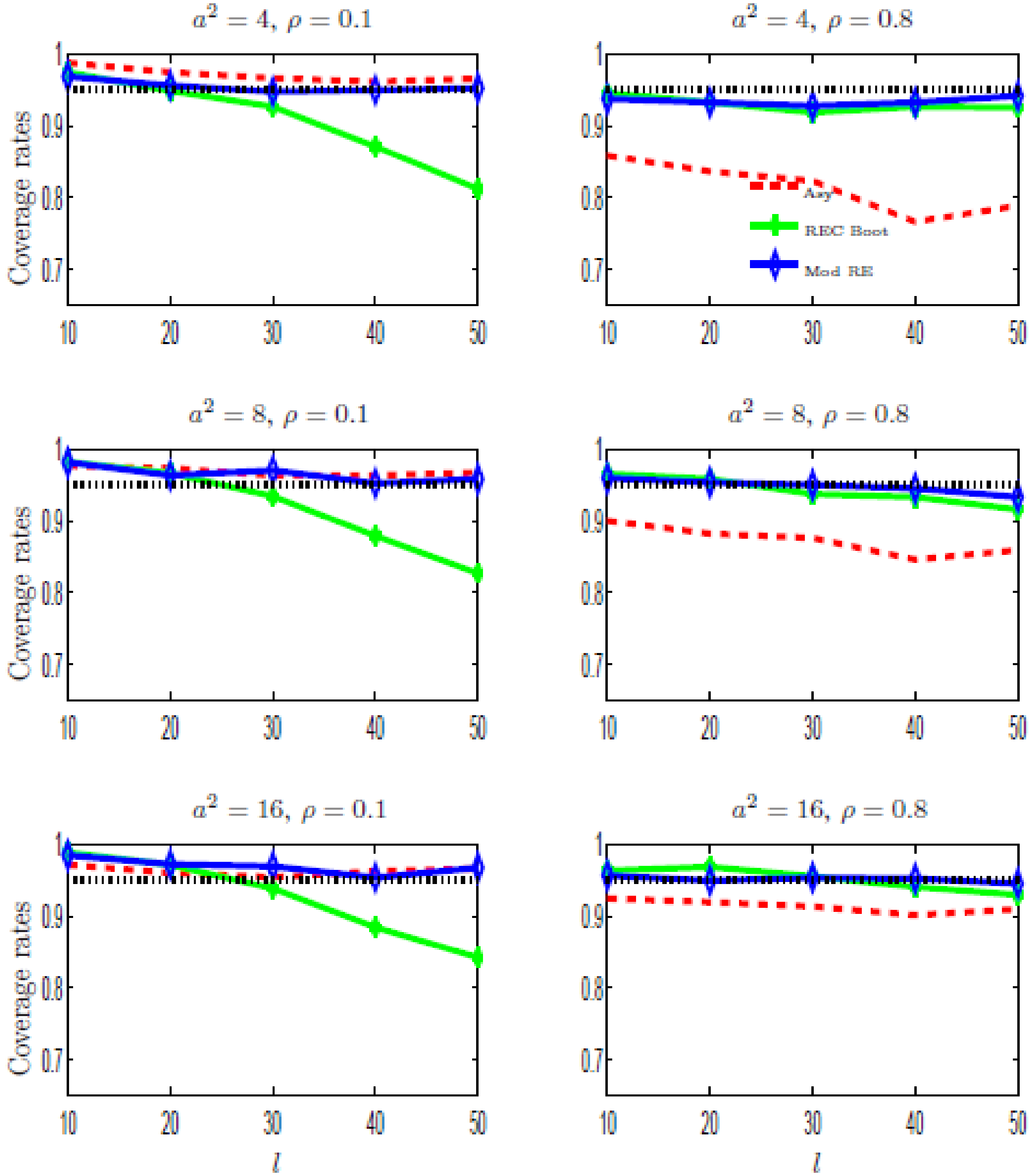

Figure 3.5: Non-rejection rates for percentile type bootstrap Wald tests as a function of l

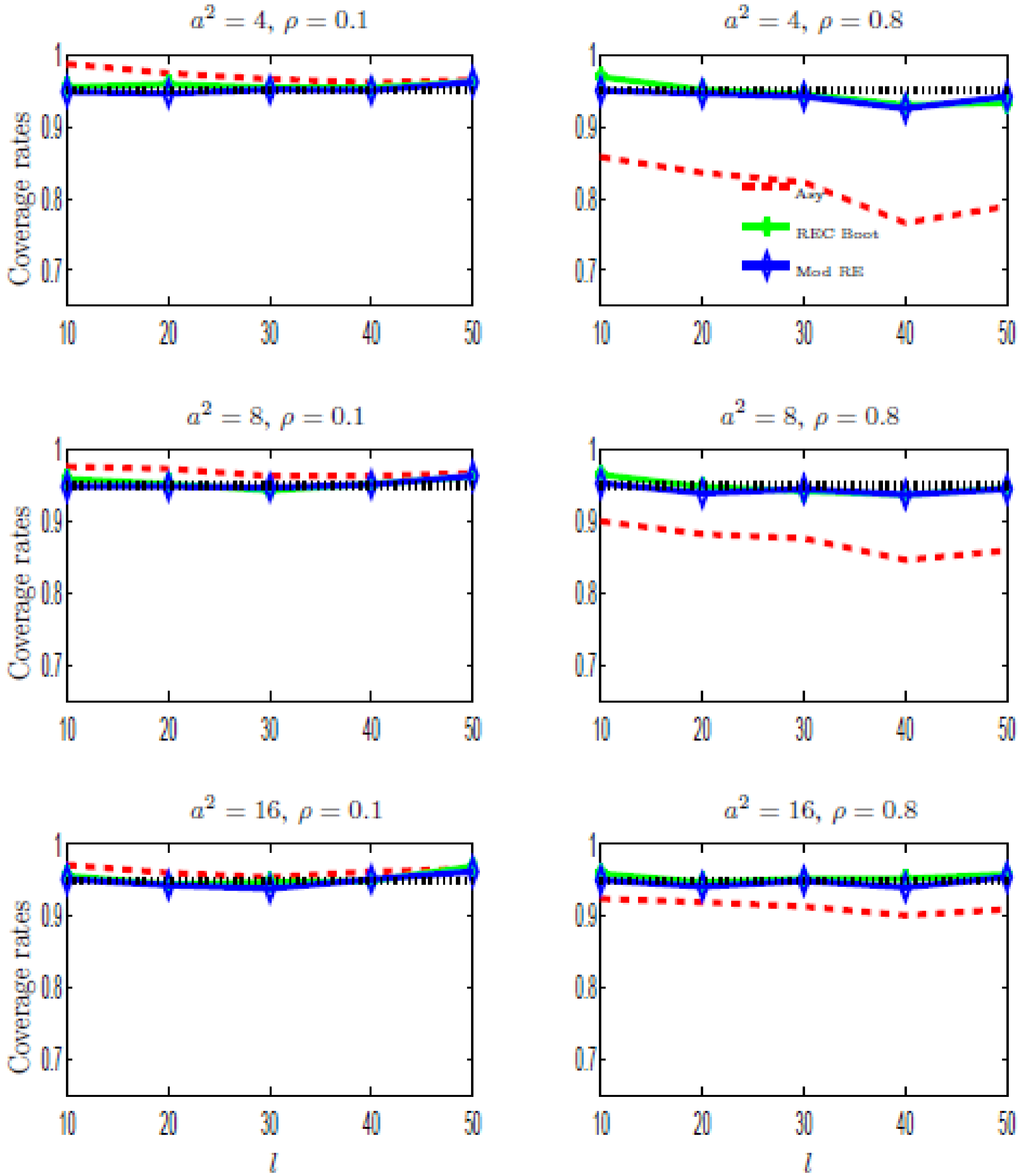

Figure 3.6: Non-rejection rates for percentile-t type bootstrap Wald tests as a function of l

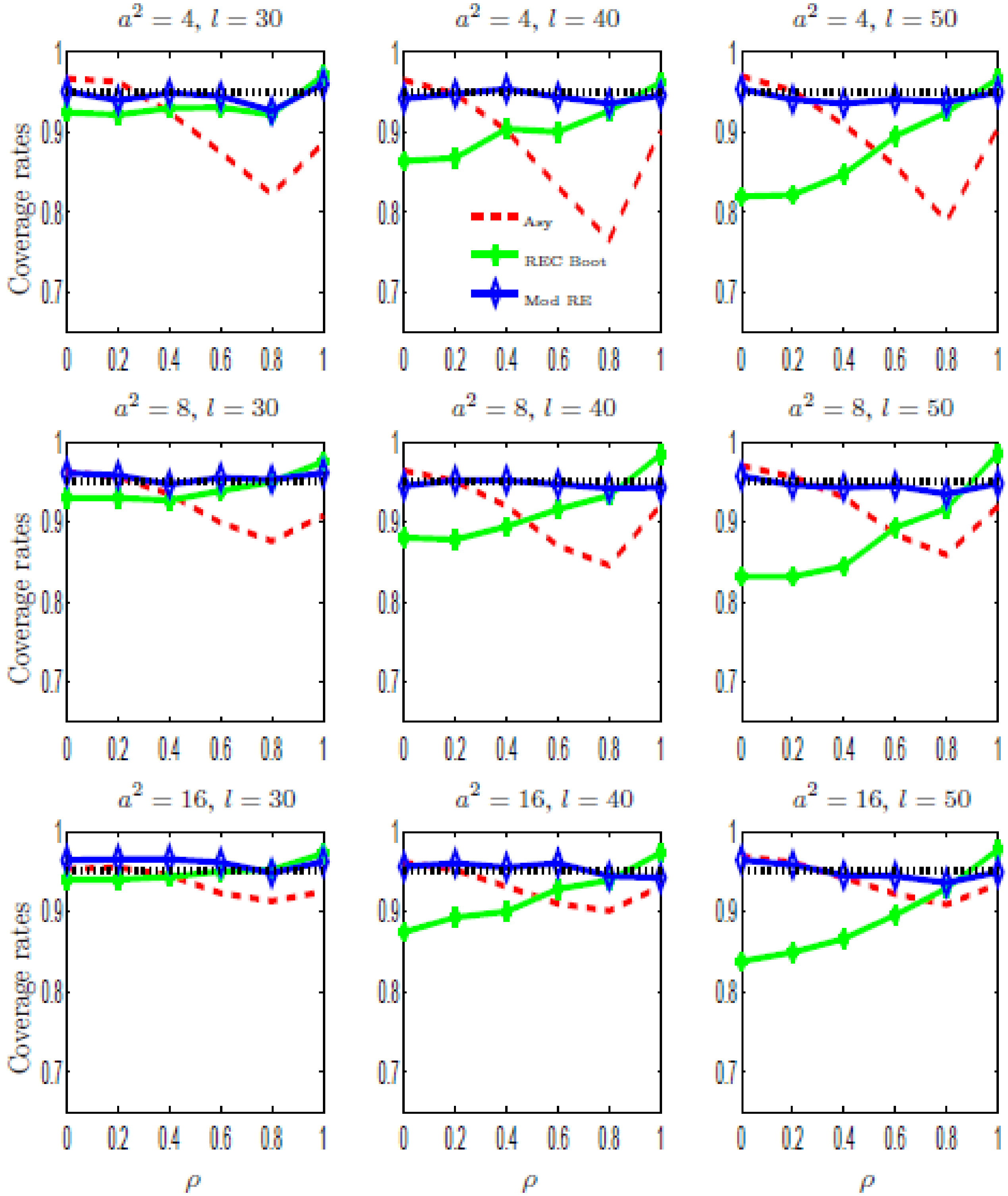

Figure 3.7: Non-rejection rates for percentile type bootstrap Wald tests as a function of ρ

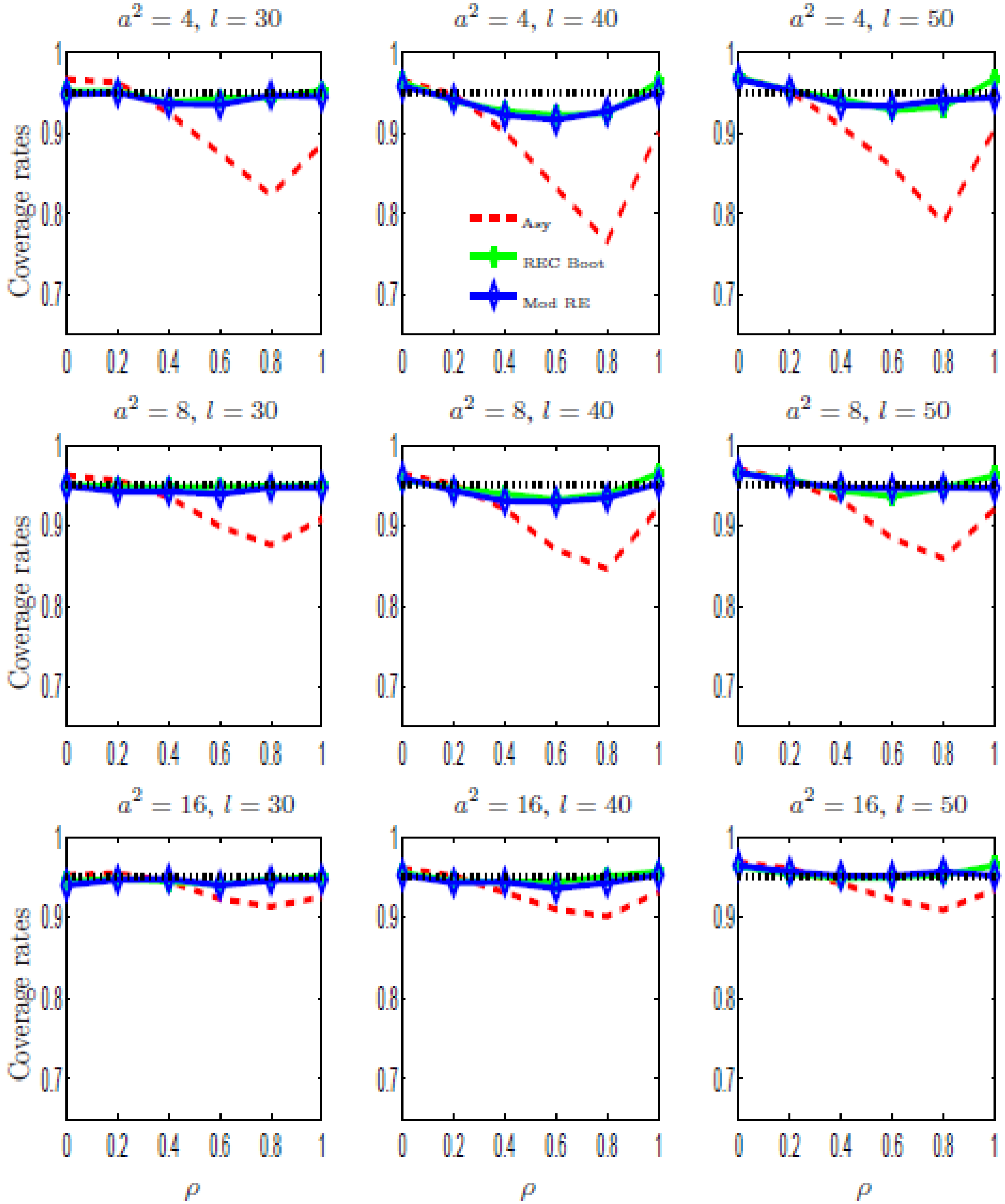

Figure 3.8: Non-rejection rates for percentile-t type bootstrap Wald tests as a function of ρ

Bibliography

Alvarez, J., and M. Arellano (2003): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," *Econometrica*, 71(4), 1121–1159.

Anatolyev, S. (2012): "Inference in regression models with many regressors," *Journal of Econometrics*, 170(2), 368–382.

Anatolyev, S., and N. Gospodinov (2011): "Specification Testing In Models With Many Instruments," *Econometric Theory*, 27(02), 427–441.

Anderson, T., N. Kunitomo, and Y. Matsushita (2010): "On the asymptotic optimality of the LIML estimator with possibly many instruments," *Journal of Econometrics*, 157(2), 191–204.

Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–58.

Angrist, J. D., and A. B. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106(4), 979–1014.

Arellano, M., and J. Hahn (2005): "Understanding Bias In Nonlinear Panel Models: Some Recent Developments," Working papers, CEMFI.

Arellano, M., and J. Hahn (2006): "A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects," *Unpublished manuscript*.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62(3), 657–81.

Beran, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83(403), 687–697.

Bickel, P. J., D. A. Freedman, P. Bickel, K. Doksum, and J. Hodges (1983): "Bootstrapping regression models with many parameters," *Festschrift for Erich L. Lehmann*, pp. 28–48.

Billingsley, P. (1995): *Probability and measure*, Wiley Series in Probability and Mathematical Statistics. 3rd ed., John Wiley & Sons Inc.

Brillinger, D. R. (1981): *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, second edition edn.

Brockwell, P. J., and R. A. Davis (1991): *Time Series: Theory and Methods*. Springer, second edition edn.

Brown, B. W., and W. K. Newey (2002): "Generalized method of moments, efficient bootstrapping, and improved inference," *Journal of Business & Economic Statistics*, 20(4), 507–517.

Bun, M. J., and M. A. Carree (2005): "Bias-Corrected Estimation in Dynamic Panel Data Models," *Journal of Business & Economic Statistics*, 23, 200–210.

Carro, J. M. (2007): "Estimating dynamic panel data discrete choice models with fixed effects," *Journal of Econometrics*, 140(2), 503–528.

Chang, Y., and J. Y. Park (2003a): "A Sieve Bootstrap For The Test Of A Unit Root," *Journal of Time Series Analysis*, 24(4), 379–400.

——— (2003b): "A Sieve Bootstrap For The Test Of A Unit Root," *Journal of Time Series Analysis*, 24(4), 379–400.

Chao, J. C., and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73(5), 1673–1692.

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): "Asymptotic Distribution Of Jive In A Heteroskedastic Iv Regression With Many Instruments," *Econometric Theory*, 28(01), 42–86.

Davidson, R., and E. Flachaire (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146(1), 162–169.

Davidson, R., and J. G. MacKinnon (2008): "Bootstrap inference in a linear equation estimated by instrumental variables," *Econometrics Journal*, 11(3), 443–477.

Davidson, R., and J. G. MacKinnon (2014): "Bootstrap confidence sets with weak instruments," *Econometric Reviews*, 33(5-6), 651–675.

Davydov, J. (1968): "Convergence of distributions generated by stationary stochastic processes," *Theory of Probability and its Applications*, pp. 691–696.

Dhaene, G., and K. Jochmans (2014): "Split-panel jackknife estimation of fixed-effect models," manuscript, Sciences Po Economics Discussion Papers.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–91.

Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1388.

Engle, R. F., N. Shephard, and K. Sheppard (2008): "Fitting vast dimensional time-varying covariance models," Discussion paper.

Everaert, G., and L. Pozzi (2007): "Bootstrap-based bias correction for dynamic panels," *Journal of Economic Dynamics and Control*, 31(4), 1160–1184.

Fernandez-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, 150(1), 71 – 85.

Fernandez-Val, I., and M. Weidner (2013): "Individual and time effects in nonlinear panel models with large N,T," CeMMAP working papers CWP60/13, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Fuller, W. A. (1977): "Some properties of a modification of the limited information estimator," *Econometrica*, 45(4), 939–53.

Galvão, A. F., and K. Kato (2013): "Estimation and Inference for Linear Panel Data Models under Misspecification when Both N and T are Large," forthcoming in the *Journal of Business and Economic Statistics*.

Goncalves, S., and H. White (2004): "Maximum likelihood and the bootstrap for nonlinear dynamic models," *Journal of Econometrics*, 119(1), 199–219.

Gonçalves, S. (2011): "The moving blocks bootstrap for panel linear regression models with individual fixed effects," *Econometric Theory*, 27, 1048–1082.

Gonçalves, S., and M. Kaffo (2013): "Bootstrap inference for linear dynamic panel data models with individual fixed effects," *Unpublished manuscript*.

Gonçalves, S., and L. Kilian (2004): "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 123(1), 89–120.

Gonçalves, S., and B. Perron (2014): "Bootstrapping Factor-Augmented Regression Models," forthcoming in the *Journal of Econometrics*.

Gonçalves, S., and H. White (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100, 970–979.

Gouriéroux, C., P. C. Phillips, and J. Yu (2010): "Indirect inference for dynamic panel models," *Journal of Econometrics*, 157(1), 68–77.

Hahn, J., and G. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both n and T Are Large," *Econometrica*, 70(4), 1639–1657.

——— (2011a): "Bias Reduction For Dynamic Nonlinear Panel Models With Fixed Effects," *Econometric Theory*, 27(06), 1152–1191.

——— (2011b): "Bias Reduction For Dynamic Nonlinear Panel Models With Fixed Effects," *Econometric Theory*, 27(06), 1152–1191.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72(4), 1295–1319.

Hall, P., and J. L. Horowitz (1996): "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica: Journal of the Econometric Society*, pp. 891–916.

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

Hansen, C. B. (2007): "Asymptotic properties of a robust variance matrix estimator for panel data when T is large," *Journal of Econometrics*, 141(2), 597–620.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3(2), 211–255.

Hounkannounon, B. (2010): "Bootstrap for panel regression models with random effects," Discussion paper.

Kaffo, M. (2013): "Bootstrap inference for nonlinear dynamic panel data models with individual fixed effects," *Unpublished manuscript*.

Kapetanios, G. (2008): "A Bootstrap Procedure for Panel Datasets with Many Cross-sectional Units," *Econometrics Journal*, 11(2), 377–395.

Kim, M. S., and Y. Sun (2013): "Bootstrap and k-step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Models," *Unpublished manuscript*.

Kiviet, J. F. (1995): "On bias, inconsistency, and efficiency of various estimators in dynamic panel data models," *Journal of Econometrics*, 68(1), 53–78.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

Kreiss, J. (1997): "Asymptotical Properties of Residual Bootstrap for Autoregressions," Manuscript, institute for mathematical stochastics, technical university of braunschweig, germany.

KUNITOMO, N. (1980): "Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations," *Journal of the American Statistical Association*, 75(371), 693–700.

LEE, Y. (2012): "Bias in dynamic panel models under time series misspecification," *Journal of Econometrics*.

LIU, R. (1988): "Bootstrap procedure under some non-i.i.d. models.," *Annals of Statistics*, 16, 1696–1708.

LIU, R., AND K. SINGH (1992): "Efficiency and robustness in resampling," *Annals of Statitics*, 20, 370–384.

MAMMEN, E. (1988): "Asymptotics with increasing dimension for robust regression with applications to the bootstrap," *The Annals of Statistics*, 17(1), pp. 382–400.

MOON, H. R., B. PERRON, AND P. C. B. PHILLIPS (2014): "Incidental Parameters and Dynamic Panel Modeling," forthcoming in handbook of panel data.

MOON, H. R., AND P. C. PHILLIPS (2000): "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," Cowles Foundation Discussion Papers 1274, Cowles Foundation for Research in Economics, Yale University.

MOON, H. R., AND P. C. B. PHILLIPS (2004): "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," *Econometrica*, 72(2), 467–522.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

Moreira, M. J., J. R. Porter, and G. A. Suarez (2009): "Bootstrap validity for the score test when instruments may be weak," *Journal of Econometrics*, 149(1), 52–64.

Morimune, K. (1983): "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared with the Sample Size," *Econometrica*, 51(3), 821–41.

Nagar, A. (1959): "The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations," *Econometrica: Journal of the Econometric Society*, pp. 575–595.

Newey, W. K., and F. Windmeijer (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77(3), 687–719.

Neyman, J., and E. Scott (1948): "Consistent estimates based on partially consistent observations," *Econometrica*, 16, 1–32.

Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49(6), 1417–26.

Pakel, C., N. Shephard, and K. Sheppard (2011): "Nuisance parameters, composite likelihoods and a panel of GARCH models," *Statistica Sinica*, 21(1), 307–329.

Phillips, P. C., and D. Sul (2007): "Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence," *Journal of Econometrics*, 137(1), 162–188.

Portnoy, S. (1984): "Asymptotic behavior of M-estimators of p regression parameters when p2/n is large. I. Consistency," *The Annals of Statistics*, pp. 1298–1309.

Rothenberg, T. J. (1984): "Approximating the distributions of econometric estimators and test statistics," *Handbook of econometrics*, 2, 881–935.

Staiger, D., and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

Stock, J. H., and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," *Identification and inference for econometric models: essays in honor of Thomas Rothenberg*.

Sun, Y., and M. S. Kim (2013): "k-step Bootstrap Bias Correction for Fixed Effects Estimators in Nonlinear Panel Models," University of california at san diego, economics working paper series, Department of Economics, UC San Diego.

Wu, C. (1986): "Jackknife, bootstrap and other resampling methods in regression analysis.," *Annals of Statistics*, 14, 1261–1295.

Xiong, S., and G. Li (2008): "Some results on the convergence of conditional distributions," *Statistics & Probability Letters*, 78(18), 3249–3253.

Annexes

.1 Proof of main results in Chapter 1

.1.1 Proofs of results in Section 1.2

Throughout this Appendix, we let Δ denote a generic constant independent of n and T. Given a matrix A, we let $|A| = (tr\,(A'A))^{1/2}$. The following results are instrumental in the proofs that follow. They correspond to Lemmas 1 and 2 in Hansen (2007) respectively.

Theorem .1.1. *Suppose Z_{iT} are independent across i for all T with $E(Z_{iT}) = \mu_{iT}$ and $E|Z_{iT}|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. Then $\frac{1}{n}\sum_{i=1}^{n}(Z_{iT} - \mu_{iT}) \to^{P} 0$ as $n, T \to \infty$ jointly.*

Theorem .1.2. *For $k \times 1$ vectors Z_{iT}, suppose Z_{iT} are independent across i for all T with $E(Z_{iT}) = 0$, $E\left(Z_{iT}Z_{iT}'\right) = \Omega_{iT}$, and $E|Z_{iT}|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$. Assume $\Omega = \lim_{n,T}\frac{1}{n}\sum_{i=1}^{n}\Omega_{iT}$ is positive definite with minimum eigenvalue $\lambda_{min} > 0$. Then $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{iT} \to^{d} N(0, \Omega)$ as $n, T \to \infty$ jointly.*

We first provide some auxiliary lemmas, followed by the proof of Theorem 1.2.1. The proof of the auxiliary lemmas follows at the end.

Lemma .1.1. *Under Assumption A1, for fixed l, $p \in \mathbb{N}$, **(i)** $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it-l}\varepsilon_{it-p} \to^{P} \sigma^{2}1_{\{l=p\}}$; and **(ii)** $\frac{1}{nT^{2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p} \to^{P} 0$.*

Lemma .1.2. *Under Assumption A1, for fixed* $k \in \mathbb{N}$, $\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} (\varepsilon_{it}\varepsilon_{it-1}, \ldots, \varepsilon_{it}\varepsilon_{it-k}) \to^{d} N(0, \Omega_k)$, *where* $\Omega_k \equiv [\tau_{lp}]_{l,p=1,\ldots,k}$.

Lemma .1.2 is the analog of Lemma A.1 of Gonçalves and Kilian (2004) (henceforth GK (2004)). To state the following lemma, we need to introduce some notation. In particular, let $u_{it} = \sum_{l=0}^{\infty} \theta_0^l \varepsilon_{it-l}$, which is well defined given that $|\theta_0| < 1$. It follows that

$$y_{it-1} = \frac{\alpha_i}{1-\theta_0} + \sum_{l=1}^{\infty} \theta_0^{l-1} \varepsilon_{it-l} \equiv \frac{\alpha_i}{1-\theta_0} + u_{it-1}, \tag{13}$$

for all (i, t). Therefore,

$$A_{nT} \equiv \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1} - \bar{y}_{i-})^2 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1}^2 - \frac{1}{n} \sum_{i=1}^{n} \bar{u}_{i-}^2 \equiv A_{nT1} - A_{nT2},$$

where $\bar{u}_{i-} = \frac{1}{T} \sum_{t=1}^{T} u_{it-1}$. The next lemma establishes the consistency of A_{nT}.

Lemma .1.3. *Under Assumption A1,* ***(i)*** $A_{nT1} \to^{P} A \equiv \sigma^2 (1-\theta_0^2)^{-1}$; ***(ii)*** $A_{nT2} \to^{P} 0$; *and* ***(iii)*** $A_{nT} \to^{P} A$.

Our next lemma establishes the limiting distribution of

$$\begin{aligned} B_{nT} &\equiv \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it-1} - \bar{y}_{i-}) (\varepsilon_{it} - \bar{\varepsilon}_i) \\ &= \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1} \varepsilon_{it} - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u_{it-1} \bar{\varepsilon}_i \equiv B_{nT1} - B_{nT2}. \end{aligned}$$

$$B_{nT} \equiv$$

Lemma .1.4. *Under Assumption A1,* ***(i)*** $B_{nT1} \to^{d} N(0, B)$, *where* $B = \sum_{l,p=1}^{\infty} \theta_0^{l+p-2} \tau_{lp}$; ***(ii)*** $B_{nT2} \to^{P} -A \cdot D$, *where* $A = \sigma^2 (1-\theta_0^2)^{-1}$ *and* $D = -\sqrt{\rho}(1+\theta_0)$; *and* ***(iii)*** $B_{nT} \to^{d} N(A \cdot D, B)$.

Proof of Theorem 1.2.1. The proof follows from Lemmas .1.3 and .1.4 by Slutsky's theorem.

Proof of Lemma .1.1 (i) For fixed l, $p \in \mathbb{N}$, let $Z_{iT}^{lp} = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it-l}\varepsilon_{it-p}$, $i = 1, \ldots, n$. We check that $\{Z_{iT}^{lp}\}$ satisfies the conditions of Theorem .1.1. First, $\{Z_{iT}^{lp}\}$ are independent across i for all T with $E\left(Z_{iT}^{lp}\right) = \sigma_i^2 1_{\{l=p\}}$. Second, we show that $E\left|Z_{iT}^{lp}\right|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i and T. Taking $\delta = 1$, by repeated application of the Cauchy-Schwartz inequality, we can show that $E\left(Z_{iT}^{lp}\right)^2 \leq E\left(\varepsilon_{it}\right)^4 \leq \Delta < \infty$. Thus, $\frac{1}{n}\sum_{i=1}^{n}\left(Z_{iT}^{lp} - \sigma_i^2 1_{\{l=p\}}\right) \to_P 0$ as $n, T \to \infty$ jointly. The result follows by noting that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 = \sigma^2$ by A1(iv). To prove part (ii), define for fixed l, $p \in \mathbb{N}$, $Z_{iT}^{lp} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{t-1} \varepsilon_{it-l}\varepsilon_{is-p}$. Then, $\{Z_{iT}^{lp}\}$ are independent across i for all T with $E\left(Z_{iT}^{lp}\right) = \mu_{iT}^{lp}$, where $\mu_{iT}^{lp} = 0$ for $l \leq p$ and for $l - p \geq T$, and $\mu_{iT}^{lp} = \frac{T-l-p}{T^2}\sigma_i^2$ for $l - p \in \{1, \ldots, T-1\}$. By repeated application of the Cauchy-Schwartz inequality, we can show that $E\left(Z_{iT}^{lp}\right)^2 \leq E\left(\varepsilon_{it}\right)^4 \leq \Delta < \infty$, which proves that Z_{iT}^{lp} verifies the conditions of Theorem .1.1. To end the proof of (ii), note that by definition of μ_{iT}^{lp},

$$\frac{1}{n}\sum_{i=1}^{n} \mu_{iT}^{lp} = \frac{T-l-p}{T^2}\left(\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2\right) 1_{\{l-p\in\{1,\ldots,T-1\}\}} \to 0$$

as $n, T \to \infty$ jointly, for all l, $p \in \mathbb{N}$, given that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \sigma_i^2 = \sigma^2$.

Proof of Lemma .1.2 For fixed $k \in \mathbb{N}$, let $Z_{iT}^{k} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\varepsilon_{it}\varepsilon_{it-1}, \ldots, \varepsilon_{it}\varepsilon_{it-k}\right)'$, $i = 1, \ldots, n$. We check that Z_{iT}^{k} satisfies the conditions of Theorem .1.2. First, Z_{iT}^{k} are independent across i for all T with $E\left(Z_{iT}^{k}\right) = 0$. Second, $E\left(Z_{iT}^{k} Z_{iT}^{k\prime}\right) = [\tau_{ilp}]_{l,p=1,\ldots,k} \equiv \Omega_{ik}$ for all i since by assumption $E\left(\varepsilon_{it}^2\varepsilon_{it-l}\varepsilon_{it-p}\right) = \tau_{ilp}$ for all t and all l, p. Third, we show that for fixed $k \in \mathbb{N}$, $E\left|Z_{iT}^{k}\right|^{2\delta} \leq \Delta <$

∞, uniformly in i for some $\delta > 1$ (we take $\delta = 2$). By the $c - r$ inequality,

$$\begin{aligned} E\left|Z_{iT}^{k}\right|^{4} &= E\left(\sum_{l=1}^{k}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{it-l}\right)^{2}\right)^{2} \leq k\sum_{l=1}^{k}E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{it-l}\right)^{4} \\ &= k\sum_{l=1}^{k}\frac{1}{T^{2}}\sum_{t_1,\ldots,t_4=1}^{T}E\left(z_{it_1}^{l}z_{it_2}^{l}z_{it_3}^{l}z_{it_4}^{l}\right), \end{aligned}$$

where we let $z_{it}^{l} = \varepsilon_{it}\varepsilon_{it-l}$ for all $1 \leq l \leq k$. Noting that $E\left(z_{it}^{l}\right) = 0$ and given the definition of the fourth order joint cumulant (see Brillinger (1981), p. 19), we have that

$$\begin{aligned} E\left(z_{it_1}^{l}z_{it_2}^{l}z_{it_3}^{l}z_{it_4}^{l}\right) &= E\left(z_{t_1}^{l}z_{t_2}^{l}\right)E\left(z_{t_3}^{l}z_{t_4}^{l}\right) + E\left(z_{t_1}^{l}z_{t_3}^{l}\right)E\left(z_{t_2}^{l}z_{t_4}^{l}\right) \\ &\quad + E\left(z_{t_1}^{l}z_{t_4}^{l}\right)E\left(z_{t_2}^{l}z_{t_3}^{l}\right) + cum\left(z_{it_1}^{l}, z_{it_2}^{l}, z_{it_3}^{l}, z_{it_4}^{l}\right). \end{aligned}$$

By the m.d.s assumption, $E\left(z_{t}^{l}z_{s}^{l}\right) = E\left(\varepsilon_{it}\varepsilon_{it-l}\varepsilon_{is}\varepsilon_{is-l}\right) = \tau_{ill}1_{\{t=s\}}$ for any (t, s), which implies that

$$\frac{1}{T^{2}}\sum_{t_1,\ldots,t_4=1}^{T}E\left(z_{it_1}^{l}z_{it_2}^{l}z_{it_3}^{l}z_{it_4}^{l}\right) = 3\tau_{ill}^{2} + \frac{1}{T^{2}}\sum_{t_1,\ldots,t_4=1}^{T}cum\left(z_{it_1}^{l}, z_{it_2}^{l}, z_{it_3}^{l}, z_{it_4}^{l}\right).$$

Given the strict stationarity assumption,

$$cum\left(z_{it_1}^{l}, z_{it_2}^{l}, z_{it_3}^{l}, z_{it_4}^{l}\right) = cum\left(z_{it_1-t_4}^{l}, z_{it_2-t_4}^{l}, z_{it_3-t_4}^{l}, z_{i0}^{l}\right),$$

which implies that

$$\frac{1}{T^{2}}\sum_{t_1,\ldots,t_4=1}^{T}E\left(z_{it_1}^{l}z_{it_2}^{l}z_{it_3}^{l}z_{it_4}^{l}\right) = 3\tau_{ill}^{2} + \frac{1}{T^{2}}\sum_{t_4=1}^{T}\left\{\sum_{t_1,t_2,t_3=1}^{T}cum\left(z_{it_1-t_4}^{l}, z_{it_2-t_4}^{l}, z_{it_3-t_4}^{l}, z_{i0}^{l}\right)\right\},$$

where the expression in curly brackets is $O(1)$ uniformly in i, l and t_4, given A1(vii) (applied with $l_1 = l_2 = l_3 = l_4 = l$). This shows that

$\frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} E\left(z_{it_1}^l z_{it_2}^l z_{it_3}^l z_{it_4}^l\right)$ is uniformly bounded in i, l and T and hence, for a fixed $k \in \mathbb{N}$, $E\left|Z_{iT}^k\right|^4 \leq \Delta < \infty$ uniformly in i and T. Also,

$$\lim_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\Omega_{ik} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left[\tau_{ilp}\right]_{l,p=1,\ldots,k} = \left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\tau_{ilp}\right]_{l,p=1,\ldots,k} = \left[\tau_{lp}\right]_{l,p=1,\ldots,k} \equiv \Omega_k,$$

where Ω_k is positive definite with minimum eigenvalue $\lambda_{min} > 0$ since by assumption, $\tau_{ll} > 0$ for all l. Thus, the conditions of Theorem .1.2 are verified, ending the proof.

Proof of Lemma .1.3. The proof of part (i) follows from Lemma .1.1(i) using the same steps as the proof that $A_{1n} \to^P 0$ in Theorem 3.1 in GK (p. 108). To prove (ii), which is new in our panel context, we use the definition of $\bar{u}_{i-}$ to decompose A_{nT2} as follows:

$$\begin{aligned} A_{nT2} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)^2 \\ &= \frac{1}{T}\left\{\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it-l}\varepsilon_{it-p}\right)\right\} \\ &\quad +2\left\{\sum_{l=1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right\} \equiv a_{1,nT} + 2a_{2,nT}. \end{aligned}$$

Given part (i), we have $a_{1,nT} = (1/T) \times A_{nT1} = o_P(1)$. Next we show that $a_{2,nT} = o_P(1)$. For fixed $m \in \mathbb{N}$, define $a_{2,nT}^m = \sum_{l=1}^{m}\sum_{p=1}^{m}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)$. By Lemma .1.1(ii), it follows that that $a_{2,nT}^m \to 0$ for all $m \in \mathbb{N}$. Thus, it suffices to show that $\lim_{m\to\infty}\limsup_{n,T\to\infty} P\left(\left|a_{2,nT} - a_{2,nT}^m\right| > \delta\right) = 0$, for all $\delta > 0$ (see Brockwell and Davis (1991)'s Proposition 6.3.9). By Markov's inequality,

$$
\begin{aligned}
P\left(\left|a_{2,nT}-a_{2,nT}^{m}\right|>\delta\right) &\leq \frac{1}{\delta}E\left|a_{2,nT}-a_{2,nT}^{m}\right| \\
&\leq \frac{1}{\delta}E\left|\sum_{l=m+1}^{\infty}\sum_{p=1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right| \\
&\quad+\frac{1}{\delta}E\left|\sum_{l=1}^{m}\sum_{p=m+1}^{\infty}\theta_0^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_{it-l}\varepsilon_{is-p}\right)\right| \\
&\leq \frac{2}{\delta}\sum_{l=m+1}^{\infty}\sum_{p=1}^{\infty}\left|\theta_0\right|^{l+p-2}\left(\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s=1}^{t-1}E\left|\varepsilon_{it-l}\varepsilon_{is-p}\right|\right) \\
&\leq \left(\sum_{l=m+1}^{\infty}\left|\theta_0\right|^{l-1}\right)K\rightarrow 0 \text{ as } m\rightarrow\infty,
\end{aligned}
$$

given the absolute summability of θ_0^{l-1} and the fact that $E\left|\varepsilon_{it-l}\varepsilon_{is-p}\right|\leq\Delta<\infty$ uniformly. This completes the proof of (ii). (iii) follows from (i) and (ii).
Proof of Lemma .1.4 Part (i) follows from Lemma .1.1 and the cross sectional independence assumption, using arguments similar to those used in the proof of Theorem 3.1 of GK (2004) (see part (ii) of their proof). To prove (ii) (which is specific to the fixed effects OLS estimator), note that we

can show that the following decomposition holds:

$$\begin{aligned}
B_{nT2} &\equiv \frac{1}{T\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{l=1}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right) \\
&= \frac{1}{T\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{l=1}^{t-1}\theta_0^{l-1}\varepsilon_{it-l}+\sum_{l=t}^{\infty}\theta_0^{l-1}\varepsilon_{it-l}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right) \\
&= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=1}^{T}\varepsilon_{is}\right) \\
&\quad +\sqrt{\frac{n}{T}}\frac{1-\theta_0^T}{1-\theta_0}\left\{\sum_{l=1}^{\infty}\theta_0^{l-1}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{i1-l}\right)\right\} \\
&\equiv \mathcal{B}_{nT2.1}+\mathcal{B}_{nT2.2}.
\end{aligned}$$

Now,

$$\begin{aligned}
\mathcal{B}_{nT2.1} &= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=1}^{T-l}\varepsilon_{is}+\sum_{s=T-l+1}^{T}\varepsilon_{is}\right) \\
&= \sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2+\sqrt{\frac{n}{T}}\sum_{l=1}^{T-1}\theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T}\varepsilon_{is}\right) \\
&\equiv b_1+b_2.
\end{aligned}$$

For fixed $m\in\mathbb{N}$, define

$$b_{1,m}=\sqrt{\frac{n}{T}}\sum_{l=1}^{m-1}\theta_0^{l-1}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2\right)=\sqrt{\frac{n}{T}}\sum_{l=1}^{m-1}\theta_0^{l-1}\left(\frac{1}{n}\sum_{i=1}^{n}Z_{iT,l}\right),$$

where $Z_{iT,l}\equiv T^{-1}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)^2$. For fixed l, we can show that $\frac{1}{n}\sum_{i=1}^{n}Z_{iT,l}\to^{P}\sigma^2$ by an application of Lemma .1.1. In particular, we can use the same arguments as in Lemma A.2 to show that $E\left|Z_{iT,l}\right|^2$ is uniformly bounded by relying on Assumption A1 (vi). Thus, $b_{1,m}\to^{P}\sqrt{\rho}\sum_{l=1}^{m-1}\theta_0^{l-1}\sigma^2=\sqrt{\rho}\sigma^2\frac{1-\theta_0^{m-1}}{1-\theta_0}\equiv D_m$ and $D_m\to\sqrt{\rho}\sigma^2\frac{1}{1-\theta_0}\equiv -A\cdot D$ as $m\to\infty$, where

$A \equiv \sigma^2/(1-\theta_0^2)$ and $D \equiv -\sqrt{\rho}\,(1+\theta_0)$. In addition, by Markov's inequality, we have

$$P(|b_1 - b_{1,m}| > \delta) \leq \frac{1}{\delta}\sqrt{\frac{n}{T}} \sum_{l=m}^{T-1} |\theta_0|^{l-1} \left(\frac{1}{n}\sum_{i=1}^{n} E(Z_{iT,l})\right)$$
$$= \frac{1}{\delta}\sqrt{\frac{n}{T}} \sum_{l=m}^{T-1} |\theta_0|^{l-1} \left(\frac{T-l}{T}\frac{1}{n}\sum_{i=1}^{n} \sigma_i^2\right).$$

It follows that $\lim_{m\to\infty}\limsup_{n,T\to\infty} P(|b_1 - b_{1,m}| > \delta) = 0$ since $n/T \to \rho$, $|\theta_0|^{l-1}$ is absolutely summable and $\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 \to \sigma^2$. Let us turn to b_2. For fixed m, define

$$b_{2,m} = \sqrt{\frac{n}{T}} \sum_{l=1}^{m-1} \theta_0^{l-1} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{T}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T}\varepsilon_{is}\right) \equiv \sqrt{\frac{n}{T}} \sum_{l=1}^{m-1} \theta_0^{l-1}\frac{1}{n}\sum_{i=1}^{n} Y_{iT,l},$$

where $Y_{iT,l}$ are independent across i, $E(Y_{iT,l}) = 0$ and $E|Y_{iT,l}|^2 \leq E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)^4 \leq \Delta$ by Assumption A1 (v) and (vi). Thus, by Theorem .1.1, $\frac{1}{n}\sum_{i=1}^{n} Y_{iT,l} = o_P(1)$ and therefore, $b_{2,m} = o_P(1)$. Finally, by Markov's inequality, we have

$$P(|b_2 - b_{2,m}| > \delta) \leq \frac{1}{\delta}\sqrt{\frac{n}{T}} E\left|\sum_{l=m}^{T-1} \theta_0^{l-1}\frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T-l}\varepsilon_{it}\right)\left(\sum_{s=T-l+1}^{T}\varepsilon_{is}\right)\right|$$
$$\leq \frac{1}{\delta}\sqrt{\frac{n}{T}}\sum_{l=m}^{T-1}|\theta_0|^{l-1}\frac{1}{n}\sum_{i=1}^{n} E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)^2 = \frac{1}{\delta}\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\right)\sum_{l=m}^{T-1}|\theta_0|^{l-1},$$

which implies that $\lim_{m\to\infty}\limsup_{n,T\to\infty} P(|b_2 - b_{2,m}| > \delta) = 0$ for any $\delta > 0$. To complete the proof of Lemma .1.4 (ii), we note that $E(\mathcal{B}_{nT2.2}) = 0$ and we can show that $Var(\mathcal{B}_{nT2.2}) = O(1/nT) = o(1)$. Part (iii) follows from (i) and (ii) by Slutsky's theorem.

.1.2 Proofs of results in Section 1.3

Proofs of results in Section 1.3.1

Throughout this section, $y_{it}^* = \hat{\alpha}_i + \hat{\theta} y_{it-1}^* + \varepsilon_{it}^*$, where $\varepsilon_{it}^* = \hat{\varepsilon}_{it} \cdot \eta_{it}$, with η_{it} are i.i.d.$(0,1)$ and $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta} y_{it-1}$.

Auxiliary lemmas

Lemma .1.5. *Under Assumption A1, for fixed $k, l \in \mathbb{N}$,* ***(i)*** $n^{-1}T^{-1}\sum_{i=1}^n \sum_{t=k+1}^T \varepsilon_{it-k}^{*2} \to^{P^*} \sigma^2$; ***(ii)*** $n^{-1}T^{-1}\sum_{i=1}^n \sum_{t=k+1}^T \varepsilon_{it-k}^* \varepsilon_{it}^* \to^{P^*} 0$; *and* ***(iii)*** $n^{-1}T^{-1}\sum_{i=1}^n \sum_{t=max(k,l)+1}^T \varepsilon_{it}^{*2} \varepsilon_{it-k}^* \varepsilon_{it-l}^* \to^{P^*} \tau_{kl} 1_{\{k=l\}}$, *in probability, where* $\tau_{kl} = E\left(\varepsilon_{it}^2 \varepsilon_{it-k} \varepsilon_{it-l}\right)$.

Lemma .1.6. *Under Assumption A1, for all $k \in \mathbb{N}$,* $\frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=k+1}^T \left(\varepsilon_{it}^* \varepsilon_{it-1}^*, \ldots, \varepsilon_{it}^* \varepsilon_{it-k}^*\right)' \to^{d^*} N\left(0, \tilde{\Omega}_k\right)$, *in probability, where* $\tilde{\Omega}_k \equiv diag\left(\tau_{11}, \ldots, \tau_{kk}\right)$.

For the next lemma, let $y_{i0}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}}$. It follows that for fixed $i = 1, \ldots, n$ and $t = 1, \ldots, T$,

$$y_{it}^* = \hat{\theta}^t \frac{\hat{\alpha}_i}{1-\hat{\theta}} + \frac{1-\hat{\theta}^t}{1-\hat{\theta}} \hat{\alpha}_i + \sum_{s=0}^{t-1} \hat{\theta}^s \varepsilon_{it-s}^* = \frac{\hat{\alpha}_i}{1-\hat{\theta}} + \sum_{s=0}^{t-1} \hat{\theta}^s \varepsilon_{it-s}^* \equiv \frac{\hat{\alpha}_i}{1-\hat{\theta}} + u_{it}^*.$$

Therefore,

$$A_{nT}^* = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \left(y_{it-1}^* - y_{i-}^*\right)^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T u_{it-1}^{*2} - \frac{1}{n} \sum_{i=1}^n \bar{u}_{i-}^{*2} \equiv A_{nT1}^* - A_{nT2}^*,$$

where $\bar{u}_{i-}^* = \frac{1}{T}\sum_{t=1}^T u_{it-1}^*$ and $u_{it-1}^* = \sum_{s=0}^{t-1-1} \hat{\theta}^s \varepsilon_{it-1-s}^* = \sum_{s=1}^{t-1} \hat{\theta}^{s-1} \varepsilon_{it-s}^*$.

Lemma .1.7. *Under Assumption A1,* ***(i)*** $A_{nT1}^* \to^{P^*} A \equiv \frac{\sigma^2}{1-\theta_0^2}$; ***(ii)*** $A_{nT2}^* \to^{P^*} 0$; *and* ***(iii)*** $A_{nT}^* \to^{P^*} A$, *in probability.*

Similarly, if we define $B_{nT}^* = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right)$, given the definition of y^*_{it-1}, we have

$$B^*_{nT} = \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u^*_{it-1} \varepsilon^*_{it} - \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} u^*_{it-1} \bar{\varepsilon}^*_{i} \equiv B^*_{nT1} - B^*_{nT2}. \quad (14)$$

Lemma .1.8. *Under Assumption A1,* ***(i)*** $B^*_{nT1} \to^{d^*} N\left(0, \tilde{B}\right)$*;* ***(ii)*** $B^*_{nT2} \to^{P^*} -A{\cdot}D$*; and* ***(iii)*** $B^*_{nT} \to^{d^*} N\left(A \cdot D, \tilde{B}\right)$*, in probability, where* $\tilde{B} = \sum_{l=1}^{\infty} \theta_0^{2l-2} \tau_{ll}$*, and* A *and* D *are defined as in Lemma .1.4.*

Proofs

Proof of Theorem 1.3.1. The result follows from Lemmas .2.3 and .2.4, Theorem .1.1 and Polya's Theorem, given that the normal distribution is everywhere continuous. Note that Assumption A1 needs to be strengthened by A1(v′) in order for $\tilde{B} = B$.

Proof of Theorem 1.3.2. We show that (1) $\sqrt{nT}(\tilde{\theta}^* - \hat{\theta}) \to^{d^*} N(D, C)$ in probability; and (2) for some $\delta > 0$, $E^*\left(\left|\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right|^{1+\delta}\right) = O_P(1)$. Starting with (1), we can write $\sqrt{nT}(\tilde{\theta}^* - \hat{\theta}) = \sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta}) + R^*_{nT}$, with $R^*_{nT} = -\sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta}) 1_{\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y^*_{it-1} - \bar{y}^*_{i-})^2 < \frac{\eta}{2}\right\}}$, given the definition of $\tilde{\theta}^*$ (with $\delta = \frac{\eta}{2}$ and $\eta \in (0, \frac{\sigma^2}{1-\theta_0^2})$). By Theorem 1.3.1, $\sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta}) = O_{P^*}(1)$, in probability, and

$$E^*\left(1_{\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y^*_{it-1} - \bar{y}^*_{i-})^2 < \frac{\eta}{2}\right\}}\right) = P^*\left(\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} (y^*_{it-1} - \bar{y}^*_{i-})^2 < \frac{\eta}{2}\right) \to^{P} 0,$$

given (1.3). By Markov's inequality, we conclude that $R^*_{nT} = o_{P^*}(1)$ in probability. To prove (2), we let $\delta = 1$ and define $S = \left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(y^*_{it-1} - \bar{y}^*_{i-})^2 \geq \frac{\eta}{2}\right\}$.

Then, given the definition of $\tilde{\theta}^*$, we have

$$E^*\left(\left|\sqrt{nT}(\tilde{\theta}^* - \hat{\theta})\right|^2\right) = E^*\left(\left|\sqrt{nT}(\hat{\theta}^*_{rd} - \hat{\theta})1_S\right|^2\right)$$
$$= E^*\left(\left(\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T(y^*_{it-1} - \bar{y}^*_{i-})^2\right)^{-2}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T(y^*_{it-1} - \bar{y}^*_{i-})(\varepsilon^*_{it} - \bar{\varepsilon}^*_i)\right)^2 1_S\right)$$
$$\leq \frac{4}{\eta^2}E^*\left(\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T(y^*_{it-1} - \bar{y}^*_{i-})(\varepsilon^*_{it} - \bar{\varepsilon}^*_i)\right)^2\right) \equiv \frac{4}{\eta^2}E^*(B^{*2}_{nT}),$$

where B^*_{nT} can be decomposed as $B^*_{nT} = B^*_{1nT} - B^*_{2nT}$, with B^*_{1nT} and B^*_{2nT} given in equation (14). We now show that $E^*(B^{*2}_{nT}) = O_P(1)$. We have that $E^*(B^{*2}_{nT}) \leq 2(E^*(B^{*2}_{1nT}) + E^*(B^{*2}_{2nT}))$, where $E^*(B^{*2}_{1nT}) = Var^*(B^*_{1nT}) \to^P \tilde{B}$, so $E^*(B^{*2}_{1nT}) = O_P(1)$. For the second term, note that

$$B^*_{2nT} = \frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T u^*_{it-1}\bar{\varepsilon}^*_i = \mathcal{B}^*_{nT2.1} + \mathcal{B}^*_{nT2.2},$$

where $\mathcal{B}^*_{nT2.1}$ and $\mathcal{B}^*_{nT2.2}$ are defined in the proof of Lemma .2.4. As we argue in that proof, $E^*(\mathcal{B}^{*2}_{nT2.2}) \to^P 0$, so we are left to prove that $E^*(\mathcal{B}^{*2}_{nT2.1}) = O_P(1)$. Given the definition of $\mathcal{B}^*_{nT2.1}$,

$$\begin{aligned}
E^*\left(\mathcal{B}^{*2}_{nT2.1}\right) &= \frac{1}{nT^3}\sum_{i,j=1}^n\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-l}\varepsilon^*_{it}\right)^2\left(\sum_{s=1}^{T-p}\varepsilon^*_{js}\right)^2\right] \\
&= \frac{1}{nT^3}\sum_{i=1}^n\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-l}\varepsilon^*_{it}\right)^2\left(\sum_{s=1}^{T-p}\varepsilon^*_{is}\right)^2\right] \\
&\quad + \frac{1}{nT^3}\sum_{i\neq j}^n\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-l}\varepsilon^*_{it}\right)^2\right]E^*\left[\left(\sum_{s=1}^{T-p}\varepsilon^*_{js}\right)^2\right] \equiv b^*_1 + b^*_2.
\end{aligned}$$

Now,

$$\begin{aligned} b_1^* &= \frac{1}{nT^3}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2l-2}E^*\left[\left(\sum_{t=1}^{T-l}\varepsilon_{it}^*\right)^4\right] + 2\frac{1}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-l}\varepsilon_{it}^*\right)^2\left(\sum_{s=1}^{T-p}\varepsilon_{is}^*\right)^2\right] \\ &= b_{11}^* + b_{12}^* \end{aligned}$$

For b_{11}^*, using the fact that $E^*|\eta_{it}|^4 \leq \Delta < \infty$,

$$\begin{aligned} b_{11}^* &\leq \frac{(1+\Delta)}{nT^3}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2l-2}\left\{\sum_{t=1}^{T-l}\hat{\varepsilon}_{it}^4 + 3\sum_{t\neq s}^{T-l}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2\right\} \\ &\leq \frac{3\,(1+\Delta)}{T}\left\{\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2\right\}\left(\sum_{l=1}^{T}\hat{\theta}^{2l-2}\right) = O_P(\frac{1}{T}), \end{aligned}$$

given that the terms in brackets are $O_P(1)$. Similarly,

$$\begin{aligned} b_{12}^* &= 2\frac{1}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-p}\varepsilon_{it}^* + \sum_{t=T-p+1}^{T-l}\varepsilon_{it}^*\right)^2\left(\sum_{s=1}^{T-p}\varepsilon_{is}^*\right)^2\right] \\ &\leq \frac{4}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}E^*\left[\left(\sum_{t=1}^{T-p}\varepsilon_{it}^*\right)^4 + \left(\sum_{t=T-p+1}^{T-l}\varepsilon_{it}^*\right)^2\left(\sum_{s=1}^{T-p}\varepsilon_{is}^*\right)^2\right] \\ &\leq \frac{4\,(1+\Delta)}{nT^3}\sum_{i=1}^{n}\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}\left\{3\sum_{t,s=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2 + \left(\sum_{t=T-p+1}^{T-l}\hat{\varepsilon}_{it}^2\right)\left(\sum_{s=1}^{T-p}\hat{\varepsilon}_{is}^2\right)\right\} \\ &\leq \frac{16\,(1+\Delta)}{T}\left\{\frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t,s=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2\right\}\left(\sum_{l>p}^{T-1}\hat{\theta}^{l+p-2}\right) = O_P(\frac{1}{T}) = O_P(1). \end{aligned}$$

Finally, for b_2^* we have

$$b_2^* = \frac{1}{nT^3}\sum_{i\neq j}^{n}\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}\left(\sum_{t=1}^{T-l}\hat{\varepsilon}_{it}^2\right)\left(\sum_{s=1}^{T-p}\hat{\varepsilon}_{js}^2\right) \leq \frac{n}{T}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\right)^2\left(\sum_{l,p=1}^{T-1}\hat{\theta}^{l+p-2}\right) = O_P(1).$$

This complete the proof of Theorem 1.3.2.

Proof of Lemma 1.3.1. From Lemma .2.3, $\hat{A}^*_{rd} \to^{P^*} A$. Hence, it suffices to show that $\hat{B}^*_{rd} \to^{P^*} \tilde{B}$, in probability. We can write $\tilde{\varepsilon}^*_{it} - \bar{\tilde{\varepsilon}}^*_i = \varepsilon^*_{it} - \bar{\varepsilon}^*_i - \left(\hat{\theta}^*_{rd} - \hat{\theta}\right)\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$, where $\tilde{\varepsilon}^*_{it} = y^*_{it} - \hat{\alpha}^*_i - \hat{\theta}^*_{rd} y^*_{it-1}$ and $\varepsilon^*_{it} = y^*_{it} - \hat{\alpha}_i - \hat{\theta} y^*_{it-1}$. Thus,

$$\hat{B}^*_{rd} = \hat{B}^*_1 + \hat{B}^*_2 + \hat{B}^*_3, \text{ with}$$

$\hat{B}^*_1 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2$, $\hat{B}^*_2 = -2\left(\hat{\theta}^*_{rd} - \hat{\theta}\right)\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 (\varepsilon^*_{it} -$ and $\hat{B}^*_3 = \left(\hat{\theta}^*_{rd} - \hat{\theta}\right)^2 \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$.

We show: (a) $\hat{B}^*_1 \to^{P^*} B$, (b) $\hat{B}^*_1 \to^{P^*} 0$ and (c) $\hat{B}^*_1 \to^{P^*} 0$. Starting with (a), note that

$$\begin{aligned}
\hat{B}^*_1 &= \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2 \\
&= \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(u^*_{it-1} - \bar{u}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2 \\
&= \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(u^{*2}_{it-1} - 2u^*_{it-1}\bar{u}^*_{i-} + \bar{u}^{*2}_{i-}\right)\left(\hat{\varepsilon}^{*2}_{it} - 2\varepsilon^*_{it}\bar{\varepsilon}^*_i + \bar{\varepsilon}^{*2}_i\right) \\
&= \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^{*2}_{it-1}\varepsilon^{*2}_{it} + R^*_{nT},
\end{aligned}$$

where

$$\begin{aligned}
R^*_{nT} &= -\frac{2}{nT}\sum_{i=1}^n \sum_{t=1}^T u^{*2}_{it-1}\varepsilon^*_{it}\bar{\varepsilon}^*_i + \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^{*2}_{it-1}\bar{\varepsilon}^{*2}_i - \frac{2}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it-1}\bar{u}^*_{i-}\varepsilon^{*2}_{it} \\
&\quad + \frac{4}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it-1}\bar{u}^*_{i-}\varepsilon^*_{it}\bar{\varepsilon}^*_i + \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \bar{u}^{*2}_{i-}\varepsilon^{*2}_{it} - \frac{3}{n}\sum_{i=1}^n \bar{u}^{*2}_{i-}\bar{\varepsilon}^{*2}_i \\
&\equiv -R^*_{nT1} + R^*_{nT2} - R^*_{nT3} + R^*_{nT4} + R^*_{nT5} - R^*_{nT6}.
\end{aligned}$$

By arguments similar to those of the proof of Corollary 3.1. of Gonçalves and Kilian (2004), one can show that $\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^{*2}_{it-1}\varepsilon^{*2}_{it} \to^{P^*} \tilde{B}$. To show that $R^*_{nT} \to^{P^*} 0$ in probability, it suffices that $E^*\left(\left|R^*_{nTj}\right|\right) \to^{P} 0$ for $j = 1, 2, 3, 5, 4, 6$. For $j = 1$,

$$|R^*_{nT1}| \leq 2\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u^{*4}_{it-1}\right]^{1/2}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} \varepsilon^{*2}_{it}\bar{\varepsilon}^{*2}_{i}\right]^{1/2} \equiv A^*_1 \times A^*_2.$$

Let us start with A^*_1. Since $u^*_{it-1} = \sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon^*_{it-s}$,

$$\begin{aligned} E^*\left|A^*_1\right|^2 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon^*_{it-s}\right)^{*4} \\ &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p,q,r=1}^{t-1}\hat{\theta}^{s+p+q+r-4}E^*\left(\varepsilon^*_{it-s}\varepsilon^*_{it-p}\varepsilon^*_{it-q}\varepsilon^*_{it-r}\right) \\ &\leq \frac{\Delta}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p=1}^{t-1}\hat{\theta}^{2s+2p-4}\hat{\varepsilon}^2_{it-s}\hat{\varepsilon}^2_{it-p} \leq \frac{\Delta}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\hat{\varepsilon}^2_{it-s}\hat{\varepsilon}^2_{it-p}, \end{aligned}$$

where $\hat{\varepsilon}_{it} = 0 \ \forall t \leq 0$. Therefore,

$$\begin{aligned} E^*\left|A^*_1\right|^2 &\leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^2_{it-s}\hat{\varepsilon}^2_{it-p}\right) \\ &\leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it-s}\right)^{1/2}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it-p}\right)^{1/2} \\ &\leq \Delta\sum_{s,p=1}^{T}\hat{\theta}^{2s+2p-4}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it}\right) = O_P(1), \end{aligned}$$

given that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}^4_{it} = O_P(1)$ under Assumption A1 and the fact that $\hat{\theta} - \theta_0 = o_P(1)$ with $|\theta_0| < 1$. To conclude that $R^*_{nT1} \to^{P^*} 0$, it suffices to show that $E^*\left(|A_2^*|\right) \rightarrow^P 0$. This can be done as follows:

$$\begin{aligned} E^*\left|A_2^*\right|^2 &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} E^*\left(\varepsilon_{it}^{*2}\bar{\varepsilon}_i^{*2}\right) = \frac{1}{nT^3}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{p,q=1}^{T} E^*\left(\varepsilon_{it}^{*2}\varepsilon_{ip}^*\varepsilon_{iq}^*\right) \\ &= \frac{1}{nT^3}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{p=1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{ip}^2 = \frac{1}{T}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\right)^2\right\} = O_P\left(\frac{1}{T}\right). \end{aligned}$$

Similar arguments can be applied to R_{nT}^*, $j = 2, 3, 5, 4, 6$. For $\hat{B}_2^*$ and $\hat{B}_3^*$, one can easily show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^3\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right)$ and $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^4$ are $O_{P^*}(1)$.

Proof of Lemma .2.1. The proof follows closely that of Lemma A.2 in GK (2004) and therefore we skip the details, only mentioning the changes introduced in the panel context. As in GK (2004), for part (i), we can write

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^{*2} - \sigma^2 = \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\left(\eta_{it}^2 - 1\right)\right] + \left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2 - \sigma^2\right] \equiv F_1^* + F_2,$$

where now $\hat{\varepsilon}_{it} = \varepsilon_{it} + (\alpha_i - \hat{\alpha}_i) + \left(\theta_0 - \hat{\theta}\right) y_{it-1}$ depends also on $(\alpha_i - \hat{\alpha}_i)$, new to the fixed effects estimator. Thus, to show that $F_2 = o_P(1)$, we need to use the fact that $\sup_{1\leq i\leq n}|\hat{\alpha}_i - \alpha_i| = o_P(1)$ under our assumptions. Since $E\left[\left(\sum_{t=1}^{T}\varepsilon_{it}\right)^2\right] = \sum_{t=1}^{T} E\left(\varepsilon_{it}^2\right) = O(T)$, it follows that $\sum_{t=1}^{T}\varepsilon_{it} = O_P\left(\sqrt{T}\right)$ uniformly in i, and therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it} = O_P(1)$ uniformly in i. Also, given that $\frac{1}{T}\sum_{t=1}^{T} y_{it-1} = O_P(1)$ uniformly in i and $\hat{\theta} - \theta_0 = o_P(1)$, we have

$$\begin{aligned}
\sup_{1\leq i\leq n}\left|\hat{\alpha}_i-\alpha_i\right| &= \sup_{1\leq i\leq n}\left|\frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)-\left(\hat{\theta}-\theta_0\right)\frac{1}{T}\sum_{t=1}^{T}y_{it-1}\right| \\
&\leq \frac{1}{\sqrt{T}}\sup_{1\leq i\leq n}\left|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}\right)\right|+\left|\hat{\theta}-\theta_0\right|\sup_{1\leq i\leq n}\left|\frac{1}{T}\sum_{t=1}^{T}y_{it-1}\right| \\
&= \frac{1}{\sqrt{T}}O_P\left(1\right)+o_P\left(1\right)O_P\left(1\right)=o_P\left(1\right).
\end{aligned}$$

The proof that $E^*\left(F_1^{*2}\right)=o_P\left(1\right)$ follows exactly the same steps as the proof in GK (2004), with the only difference that we again rely on the uniform convergence (over i) of $\hat{\alpha}_i$ towards α_i (in addition to the convergence of $\hat{\theta}$ towards θ_0) to show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^4=O_P\left(1\right)$. The proof of (ii) and (iii) follow similarly. In particular, to prove (iii) we show that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=max(k,l)+1}^{T}\varepsilon_{it}^2\varepsilon_{it-k}\varepsilon_{it-l}\to^P\tau_{kl}$ by verifing the conditions of Theorem .1.1.

Proof of Lemma .2.2. For fixed $k\in\mathbb{N}$, we check that $Z_{iT}^{*k}=\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\left(\varepsilon_{it}^*\varepsilon_{it-1}^*,\ldots,\varepsilon_{it}^*\varepsilon_{it-k}^*\right)'$ satisfies the conditions of Theorem .1.2, conditionally on the original sample with probability converging to one. First, $\{Z_{iT}^{*k}\}$ are (conditionally) independent across i for all T with $E^*\left(Z_{iT}^{*k}\right)=0$. Second,

$$E^*\left(Z_{iT}^{*k}Z_{iT}^{*k'}\right)=diag\left(\frac{1}{T}\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{it-1}^2,\ldots,\frac{1}{T}\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{it-k}^2\right)\equiv\hat{\Omega}_{iT}.$$

Under our assumptions, $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{it-p}^2\to^P\tau_{pp}$, $p=1,\ldots,k$, which implies that $plim_{n,T\to\infty}\frac{1}{n}\sum_{i=1}^{n}\hat{\Omega}_{iT}=\tilde{\Omega}_k$, where $\tilde{\Omega}_k$ is positive definite with minimum eigenvalue $\lambda_{min}>0$ since $\tau_{rr}>0$ for all $r\geq 1$. Lastly, we can show that $E^*\left\|Z_{iT}^{*k}\right\|^{2\delta}=O_P\left(1\right)$, uniformly in i for $\delta=2$. In particular, by the $c-r$ inequality (with $r=2$),

$$E^{*}\left\|Z_{iT}^{*k}\right\|^{4}=E^{*}\left(\sum_{l=1}^{k}\left(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\varepsilon_{it}^{*}\varepsilon_{it-l}^{*}\right)^{2}\right)^{2}\leq k^{2-1}\sum_{l=1}^{k}E^{*}\left(\frac{1}{\sqrt{T}}\sum_{t=k+1}^{T}\varepsilon_{it}^{*}\varepsilon_{it-l}^{*}\right)^{4}$$

$$=k\sum_{l=1}^{k}\frac{1}{T^{2}}\sum_{t_{1},\ldots,t_{4}=k+1}^{T}E^{*}\left(\varepsilon_{it_{1}}^{*}\varepsilon_{it_{1}-l}^{*}\varepsilon_{it_{2}}^{*}\varepsilon_{it_{2}-l}^{*}\varepsilon_{it_{3}}^{*}\varepsilon_{it_{3}-l}^{*}\varepsilon_{it_{4}}^{*}\varepsilon_{it_{4}-l}^{*}\right)$$

$$\leq\Delta k\sum_{l=1}^{k}\frac{1}{T^{2}}\left(\sum_{t=k+1}^{T}\hat{\varepsilon}_{it}^{4}\hat{\varepsilon}_{it-l}^{4}+3\sum_{t\neq s}^{T}\hat{\varepsilon}_{it}^{2}\hat{\varepsilon}_{it-l}^{2}\hat{\varepsilon}_{is}^{2}\hat{\varepsilon}_{is-l}^{2}\right)=O_{P}\left(1\right),$$

given that $1/T\sum_{t=1+k}^{T}\hat{\varepsilon}_{it}^{4}\hat{\varepsilon}_{it-l}^{4}=O_{P}\left(1\right)$ under Assumption A1. Note also the use of the definition of $\varepsilon_{it}^{*}=\hat{\varepsilon}_{it}\eta_{it}$ and the i.i.d. properties of η_{it} to justify the fact that the only non-zero contributions to the sum in the second equality are when (1) $t_1=t_2=t_3=t_4$; (2) $t_1=t_2\neq t_3=t_4$; (3) $t_1=t_3\neq t_2=t_4$; (4) $t_1=t_4\neq t_2=t_3$.

Proof of Lemma .2.3 The proof of (i) follows the same arguments of the proof of Lemma A.4 of GK (2004), by replacing their Lemma A.2 with our Lemma .2.1 to justify the convergence in probability of $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T}\varepsilon_{it-k}^{*2}$ towards σ^2 and of $n^{-1}T^{-1}\sum_{i=1}^{n}\sum_{t=k+1}^{T-l}\varepsilon_{it-k}^{*}\varepsilon_{it}^{*}$ towards zero. Part (iii) follows from (i) and (ii). Part (ii) is new to the panel context considered here, so we provide more details. First, recall that $u_{it-1}^{*}=\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}$, which implies that

$$\bar{u}_{i-}^{*}\equiv\frac{1}{T}\sum_{t=1}^{T}u_{it-1}^{*}=\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\hat{\theta}^{s-1}\varepsilon_{it-s}^{*}\right)=\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\underbrace{\left(\frac{1}{T}\sum_{t=1}^{T-l}\varepsilon_{it}^{*}\right)}_{\equiv\chi_{il}^{*}}=\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\chi_{il}^{*}.$$

Hence,

$$A_{nT2}^{*}=\frac{1}{n}\sum_{i=1}^{n}\bar{u}_{i-}^{*2}=\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l=1}^{T-1}\hat{\theta}^{l-1}\chi_{il}^{*}\right)^{2}=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\chi_{il}^{*2}$$
$$+\frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l-1)}\hat{\theta}^{(k-1)}\chi_{il}^{*}\chi_{il+k}^{*}\equiv\mathcal{A}_{1}^{*}+\mathcal{A}_{2}^{*}.$$

Given the definition of χ_{il}^{*}, we have that

$$\mathcal{A}_{1}^{*}=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T^{2}}\sum_{t=1}^{T-l}\varepsilon_{it}^{*2}+2\frac{1}{T^{2}}\sum_{k=1}^{T-l-1}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)\equiv a_{11}^{*}+a_{12}^{*}.$$

Using Lemma .2.1.(i), and following the proof of Lemma A.4 of GK(2004), we can show that

$$a_{11}^{*}=\frac{1}{T}\left\{\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l}\varepsilon_{it}^{*2}\right)\right\}=O_{P^{*}}\left(\frac{1}{T}\right)=o_{P^{*}}\left(1\right).$$

For the second term, we have that

$$a_{12}^{*}=\frac{2}{T}\sum_{l=1}^{T-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-l-1}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right)=\frac{2}{T}\sum_{l=1}^{T-1}\sum_{k=1}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right).$$

For fixed m, let

$$a_{12,m}^{*}=\frac{2}{T}\sum_{l=1}^{m-1}\sum_{k=1}^{m-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\right).$$

By Lemma .2.1.(ii), we have that $\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*}\varepsilon_{it+k}^{*}\to^{P^{*}}0$, in probability. Since $\hat{\theta}\to^{P}\theta_{0}$, it follows that $a_{12,m}^{*}\to^{P^{*}}0$, in probability. To conclude that $a_{12}^{*}\to^{P^{*}}0$, in probability, it suffices to show that $\lim_{m\to\infty}\limsup_{n,T\to\infty}P^{*}\left(\left|a_{12}^{*}-a_{12,m}^{*}\right|>\delta\right)=$

$o_P(1)$. We have that

$$a_{12}^* - a_{12,m}^* = \frac{2}{T}\sum_{l=m}^{T-1}\sum_{k=1}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^*\varepsilon_{it+k}^*\right)$$
$$+\frac{2}{T}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\hat{\theta}^{2(l-1)}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^*\varepsilon_{it+k}^*\right) = R_{12.1,m}^* + R_{12.2,m}^*.$$

By the triangle inequality,

$$E^*\left|R_{12.1,m}^*\right| \le \frac{2}{T}\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\sum_{k=1}^{T-l-1}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}E^*\left|\varepsilon_{it}^*\varepsilon_{it+k}^*\right|\right)$$
$$\le \frac{2}{T}\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\sum_{k=1}^{T-l-1}\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\left|\hat{\varepsilon}_{it}\hat{\varepsilon}_{it+k}\right|E^*\left|\eta_{it}\eta_{it+k}\right|\right) \le 2\Delta\left(\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\right)\left(\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)}\right),$$

where we have used the fact that $E^*\left|\eta_{it}\eta_{it+k}\right| \le \Delta$ and Cauchy-Schwartz's inequality to justify the third inequality. Under Assumption A1, we have that $\frac{1}{T}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2 = O_P(1)$ whereas $\sum_{l=m}^{T-1}\hat{\theta}^{2(l-1)} \to^P \theta_0^{2(m-1)}/\left(1-\theta_0^2\right)$, which converges to 0 as $m\to\infty$ since $|\theta_0|<1$. This shows that $\lim_{m\to\infty}\limsup_{n,T\to\infty}E^*\left|R_{12.1,m}^*\right| = o_P(1)$. For $R_{12.2,m}^*$,

$$E^*\left|R_{12.2,m}^*\right|^2 \le \frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\sum_{p=1}^{m-1}\sum_{q=m-p}^{T-p-1}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\sum_{j=1}^{n}\sum_{s=1}^{T-p-q}E^*\left(\varepsilon_{it}^*\varepsilon_{it+k}^*\varepsilon_{js}^*\varepsilon_{js+q}^*\right)\right)$$
$$=\frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{k=m-l}^{T-l-1}\sum_{p=1}^{m-1}\sum_{q=m-p}^{T-p-1}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-q)}E^*\left(\varepsilon_{it}^{*2}\varepsilon_{it+k}^*\varepsilon_{it+q}^*\right)\right)$$
$$=\frac{4}{T^2}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\sum_{k=\max(m-l,m-p)}^{\min(T-l-1,T-p-1)}\hat{\theta}^{2(l+p-2)}\left(\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-q)}E^*\left(\varepsilon_{it}^{*2}\varepsilon_{it+k}^{*2}\right)\right)$$
$$\le 4\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^4\right)\frac{1}{nT^2}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\hat{\theta}^{2(l+p-2)},$$

which converges to 0 as $n, T \to \infty$ since under Assumption 1, we have that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{4} = O_P(1)$ and

$$p\lim_{n,T\to\infty}\sum_{l=1}^{m-1}\sum_{p=1}^{m-1}\hat{\theta}^{2(l+p-2)} = \left(\frac{1-\theta_0^{2(m-1)}}{1-\theta_0^2}\right)^2 \to \frac{1}{1-\theta_0^2} \text{ as } m \to \infty,$$

showing that $\lim_{m\to\infty}\limsup_{n,T\to\infty} E^*\left|R_{12.2,m}^*\right|^2 = o_P(1)$. This ends the proof of $\mathcal{A}_1^* = o_{P^*}(1)$. For $\mathcal{A}_2^*$, we have that

$$\begin{aligned}\mathcal{A}_2^* &= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l}\varepsilon_{it}^*\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon_{is}^*\right)\\
&= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l-k}\varepsilon_{it}^* + \frac{1}{T}\sum_{t=T-l-k+1}^{T-l}\varepsilon_{it}^*\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon_{is}^*\right)\\
&= \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=1}^{T-l-k}\varepsilon_{it}^*\right)^2\\
&+ \frac{2}{n}\sum_{i=1}^{n}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{T}\sum_{t=T-l-k+1}^{T-l}\varepsilon_{it}^*\right)\left(\frac{1}{T}\sum_{s=1}^{T-l-k}\varepsilon_{is}^*\right) \equiv a_{21}^* + a_{22}^*.\end{aligned}$$

Now,

$$\begin{aligned}a_{21}^* &= \frac{2}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*2}\right)\\
&+ \frac{4}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\sum_{p=1}^{T-l-k-1}\hat{\theta}^{(l+k-2)}\left(\frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k-p}\varepsilon_{it}^*\varepsilon_{it+p}^*\right) \equiv a_{21.1}^* + a_{21.2}^*.\end{aligned}$$

By Lemma .2.1 (i), and following the proof of Lemma A.4 of GK(2004), we can show that

$$a_{21.1}^* = \frac{2}{T}\sum_{k=1}^{T-2}\sum_{l=1}^{T-1-k}\hat{\theta}^{(l+k-2)}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T-l-k}\varepsilon_{it}^{*2}\right) = O_{P^*}\left(\frac{1}{T}\right) = o_{P^*}(1).$$

The proof that $a^*_{21.2} = o_{P^*}(1)$ follows by showing that $E^* \left|a^*_{21.2}\right|^2 = o_P(1)$. For a^*_{22}, we use Markov's inequality and apply the same reasoning as that used to show that $b_2 = o_P(1)$ in the proof of Lemma A.4.

Proof of Lemma .2.4. Part (i) follows by the same arguments used by GK (2004) to prove their Lemma A.5, given our Assumption A1 and the fact that $\sup_i |\hat{\alpha}_i - \alpha_i| = o_P(1)$ and $\hat{\theta} \to^P \theta_0$. Part (iii) follows trivially from parts (i) and (ii). Part (ii) is the new bias term, which we consider in more detail here. First, recall that $\sum_{t=1}^T u^*_{it-1} = \left(\sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon^*_{it}\right)$, which implies that

$$B^*_{nT2} = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T u^*_{it-1} \bar{\varepsilon}^*_i = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \left(\sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon^*_{it}\right)\left(T^{-1} \sum_{s=1}^T \varepsilon^*_{is}\right),$$

given the definition of $\bar{\varepsilon}^*_i$. It follows that

$$\begin{aligned} B^*_{nT2} &= \frac{1}{T\sqrt{nT}} \sum_{i=1}^n \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon^*_{it}\right)\left(\sum_{s=1}^{T-l} \varepsilon^*_{is} + \sum_{s=T-l+1}^{T-l} \varepsilon^*_{is}\right) \\ &= \frac{1}{T\sqrt{nT}} \sum_{i=1}^n \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon^*_{it}\right)^2 + \frac{1}{T\sqrt{nT}} \sum_{i=1}^n \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \left(\sum_{t=1}^{T-l} \varepsilon^*_{it}\right)\left(\sum_{s=T-l+1}^{T-l} \varepsilon^*_{is}\right) \\ &\equiv \mathcal{B}^*_{nT2.1} + \mathcal{B}^*_{nT2.2}. \end{aligned}$$

For fixed l, we can write

$$\left(\sum_{t=1}^{T-l} \varepsilon^*_{it}\right)^2 = \sum_{t=1}^{T-l} \varepsilon^{*2}_{it} + 2 \sum_{k=1}^{T-l-1} \sum_{t=1}^{T-l-k} \varepsilon^*_{it} \varepsilon^*_{it+k},$$

which implies that

$$\mathcal{B}^*_{nT2.1} = \frac{1}{T\sqrt{nT}} \sum_{i=1}^n \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{t=1}^{T-l} \varepsilon^{*2}_{it} + \frac{2}{T\sqrt{nT}} \sum_{i=1}^n \sum_{l=1}^{T-1} \hat{\theta}^{l-1} \sum_{k=1}^{T-l-1} \sum_{t=1}^{T-l-k} \varepsilon^*_{it} \varepsilon^*_{it+k} \equiv b^*_1 + b^*_2.$$

Using arguments similar to those applied in the proof of Lemma .2.3, we can show that $b_1^* \to^{P^*} \sqrt{\rho}\frac{\sigma^2}{1-\theta_0}$. For b_2^*, we have that

$$\begin{aligned}
E^*\left|b_2^*\right|^2 &= 4\left(\frac{n}{T}\right)\sum_{l=1}^{T-1}\sum_{k=1}^{T-l-1}\sum_{p=1}^{T-1}\sum_{q=1}^{T-p-1}\hat{\theta}^{l+p-2}\frac{1}{n^2T^2}\sum_{i,j=1}^{n}\sum_{t=1}^{T-l-k}\sum_{s=1}^{T-p-q}E^*\left(\varepsilon_{it}^*\varepsilon_{it+k}^*\varepsilon_{js}^*\varepsilon_{js+q}^*\right)\\
&= 4\left(\frac{n}{T}\right)\sum_{l=1}^{T-1}\sum_{k=1}^{T-l-1}\sum_{p=1}^{T-1}\sum_{q=1}^{T-p-1}\hat{\theta}^{l+p-2}\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-q)}E^*\left(\varepsilon_{it}^{*2}\varepsilon_{it+k}^*\varepsilon_{it+q}^*\right)\\
&= 4\left(\frac{n}{T}\right)\sum_{l=1}^{T-1}\sum_{k=1}^{\min(T-l-1,T-p-1)}\sum_{p=1}^{T-1}\hat{\theta}^{l+p-2}\frac{1}{n^2T^2}\sum_{i=1}^{n}\sum_{t=1}^{\min(T-l-k,T-p-k)}\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{it+k}^2\\
&\leq 4\frac{1}{nT}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^4\right)\left(\frac{n}{T}\right)T\sum_{l=1}^{T-1}\sum_{p=1}^{T-1}\hat{\theta}^{l+p-2} = O_P\left(\frac{1}{n}\right).
\end{aligned}$$

Using similar arguments, we can show that $E^*\left|\mathcal{B}_{nT2.2}^*\right|^2 \to^P 0$, which completes the proof of Lemma .2.4.

Proofs of results in Section 1.3.2

Proof of Theorem 1.3.3.

We show that

$$B_{nT}^* \equiv \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}-\bar{y}_{i-}\right)\left(\varepsilon_{it}^*-\bar{\varepsilon}_i^*\right) \to^{d^*} N\left(0,B\right)$$

in probability, where $\varepsilon_{it}^* = \hat{\varepsilon}_{it}\cdot\eta_{it}$, with η_{it} are i.i.d.$(0,1)$. We can write

$$B_{nT}^* = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}\varepsilon_{it}^* - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it-1}\bar{\varepsilon}_i^* \equiv B_{nT1}^* - B_{nT2}^*.$$

Writing $B_{nT1}^* = n^{-1/2}\sum_{i=1}^{n}Z_{iT}^*$, with $Z_{iT}^* \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it-1}\varepsilon_{it}^*$, we verify that the conditions of Theorem .1.2 hold with probability converging to one. First, $\{Z_{iT}^*\}$ are independent across i for all T with $E^*\left(Z_{iT}^*\right)=0$ and $E^*\left(Z_{iT}^{*2}\right) = \frac{1}{T}\sum_{t=1}^{T}u_{it-1}^2\hat{\varepsilon}_{it}^2 \equiv \Omega_{iT}$. Moreover, for $\delta = 2$, and using the independence of

ε_{it}^* across i and t, we have that

$$E^*\left(Z_{iT}^{*2+\delta}\right) = E^*\left(\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it-1}\varepsilon_{it}^*\right)^4\right) = \frac{1}{T^2}\sum_{t,s,p,q=1}^{T} u_{it-1}u_{is-1}u_{ip-1}u_{iq-1}E^*\left(\varepsilon_{it}^*\varepsilon_{is}^*\varepsilon_{ip}^*\varepsilon_{iq}^*\right)$$
$$\leq \frac{3}{T^2}\sum_{t,s=1}^{T} u_{it-1}^2 u_{is-1}^2 E^*\left(\varepsilon_{it}^{*2}\varepsilon_{is}^{*2}\right) \leq \frac{3\Delta}{T^2}\sum_{t=1}^{T} u_{it-1}^4\hat{\varepsilon}_{it}^4 + \frac{6\Delta}{T^2}\sum_{t>s=1}^{T} u_{it-1}^2 u_{is-1}^2\hat{\varepsilon}_{it}^2\hat{\varepsilon}_{is}^2 = O_P(1)$$

given that $E^*\left|\eta_{it}\right|^4 \leq \Delta < \infty$. Finally, we can show that $\frac{1}{n}\sum_{i=1}^{n}\Omega_{iT} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it-1}^2\hat{\varepsilon}_{it}^2 \to^P B$. To complete the proof, we show that $B_{nT2}^* = \sqrt{\frac{n}{T}}\frac{1}{n}\sum_{i=1}^{n} Z_{iT}^* \to^{P^*} 0$ by verifying that the conditions of Theorem .1.1 apply to

$$Z_{iT}^* \equiv \frac{1}{T}\sum_{t=1}^{T} u_{it-1}\bar{\varepsilon}_i^* = \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it-1}\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}^*\right).$$

Given that ε_{is}^* are independent across i, so are Z_{iT}^*. Moreover, $E^*\left(Z_{iT}^*\right) = 0$ and for $\delta = 1$,

$$\begin{aligned}
E^*\left(Z_{iT}^{*1+\delta}\right) &= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it-1}\right)^2 E^*\left(\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}^*\right)^2\right) \\
&= \left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it-1}\right)^2\left(\frac{1}{T}\sum_{s=1}^{T}\hat{\varepsilon}_{is}^2\right) = O_P(1).
\end{aligned}$$

Proofs of results in Section 1.3.3

Proof of Theorem 1.3.4.

Let $I_1, \ldots, I_n$ be i.i.d. random variables uniformly distributed on $\{1, \ldots, n\}$, and let

$$\left(y_{it}^*, y_{it-1}^*\right) = \left(y_{I_i t}, y_{I_i t-1}\right),\ t = 1, \ldots, T,\ i = 1, \ldots, n.$$

Define $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \hat{\theta}y_{it-1}$, $\hat{\varepsilon}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\theta}y_{it-1}^*$ and $\varepsilon_{it}^* = y_{it}^* - \alpha_i^* - \theta_0 y_{it-1}^*$, where $\alpha_i^* = \alpha_{I_i}$. We show that (a) $A_{nT}^* \equiv \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2 \to^{P^*} A$ and (b) $B_{nT}^* \equiv \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\hat{\varepsilon}_{it}^* - \bar{\hat{\varepsilon}}_i^*\right) \to^{d^*} N(0, B)$, in probability. Recall that $y_{it-1} = \frac{\alpha_i}{1-\theta_0} + u_{it-1}$. Similarly, define $\mu_i \equiv E\left(y_{it-1}\right) = \frac{\alpha_i}{1-\theta_0}$ and $\mu_i^* = \mu_{I_i}$. Then, for (a),

$$\begin{aligned} A_{nT}^* &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left(y_{it-1}^* - \mu_i^* - \bar{y}_{i-}^* + \mu_i^*\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left(\left(y_{it-1}^* - \mu_i^*\right) - \left(\frac{1}{T}\sum_{s=1}^T\left(y_{is-1}^* - \mu_i^*\right)\right)\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\left(y_{it-1}^* - \mu_i^*\right)^2 - \frac{1}{n}\sum_{i=1}^n\left(\frac{1}{T}\sum_{s=1}^T\left(y_{is-1}^* - \mu_i^*\right)\right)^2 \equiv A_{nT1}^* - A_{nT2}^*. \end{aligned}$$

We show that (a1) $A_{nT1}^* \to^{P^*} A$ and (a2) $A_{nT2}^* \to^{P^*} 0$. For (a1), we let $Z_{iT}^* = \frac{1}{T}\sum_{t=1}^T\left(y_{it-1}^* - \mu_i^*\right)^2$, which implies that $A_{nT1}^* = \frac{1}{n}\sum_{i=1}^n Z_{iT}^*$, and we use Theorem .1.1. Notice that $\{Z_{iT}^*\}$ are independent across i for all T with

$$E^*\left(Z_{iT}^*\right) = \frac{1}{T}\sum_{t=1}^T E^*\left(\left(y_{I_it-1} - \mu_{I_i}\right)^2\right) = \frac{1}{T}\sum_{t=1}^T\frac{1}{n}\sum_{j=1}^n\left(y_{jt-1} - \mu_j\right)^2 = \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it-1}^2 \to^P A.$$

Also, for $\delta = 1$,

$$E^*\left(Z_{iT}^{*1+\delta}\right) = \frac{1}{T^2}\sum_{t,s=1}^T E^*\left(\left(y_{it-1}^* - \mu_i^*\right)^2\left(y_{is-1}^* - \mu_i^*\right)^2\right) = \frac{1}{nT^2}\sum_{i=1}^n\sum_{t,s=1}^T u_{it-1}^2 u_{is-1}^2 = O_P\left(1\right).$$

For (a2), define $\tilde{Z}_{iT}^* = \left(\frac{1}{T}\sum_{t=1}^T\left(y_{it-1}^* - \mu_i^*\right)\right)^2$ and let $A_{nT2}^* = \frac{1}{n}\sum_{i=1}^n \tilde{Z}_{iT}^*$, where the $\{\tilde{Z}_{iT}^*\}$ are independent across i for all T with

$$E^*\left(\tilde{Z}_{iT}^*\right) = \frac{1}{T^2}\sum_{t,s=1}^T E^*\left(\left(y_{it-1}^* - \mu_i^*\right)\left(y_{is-1}^* - \mu_i^*\right)\right) = \frac{1}{nT^2}\sum_{i=1}^n\sum_{t,s=1}^T u_{it-1}u_{is-1} = \frac{1}{n}\sum_{i=1}^n \bar{u}_{i-}^2 \to^P 0,$$

by Lemma .1.3. The result follows by showing that

$$E^*\left(\tilde{Z}_{iT}^{*1+\delta}\right) = \frac{1}{nT^4}\sum_{i=1}^{n}\sum_{t,s,p,q=1}^{T} u_{it-1}u_{is-1}u_{ip-1}u_{iq-1} = O_P(1)$$

for $\delta = 1$. Next we show (b). With our notations, B_{nT}^* can be rewritten as

$$B_{nT}^* = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right) - \frac{1}{nT}\sum_{j=1}^{n}\sum_{s=1}^{T}\left(y_{js-1} - \bar{y}_{j-}\right)\left(\varepsilon_{js} - \bar{\varepsilon}_j\right)\right\}$$

$$+\left\{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1} - \bar{y}_{i-}\right)^2 - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)^2\right\}\sqrt{nT}\left(\hat{\theta} - \theta_0\right) \equiv B_{nT}^{*'} + R_{nT}^*.$$

Using (a) and Theorem 1.2.1, we have $R_{nT}^* = o_{P^*}(1)\, O_P(1) = o_{P^*}(1)$. Therefore, (b) follows if we prove that $B_{nT}^{*'} \to_{d_{P^*}} N(0, B)$ in probability. Noting that $\sum_{t=1}^{T}\left(y_{it-1}^* - \bar{y}_{i-}^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right) = \sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right)$,

$$\begin{aligned}
B_{nT}^{*'} &= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(y_{it-1}^* - \mu_i^*\right)\left(\varepsilon_{it}^* - \bar{\varepsilon}_i^*\right) - \frac{1}{nT}\sum_{j=1}^{n}\sum_{s=1}^{T} u_{js-1}\left(\varepsilon_{js} - \bar{\varepsilon}_j\right)\right\} \\
&= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(y_{it-1}^* - \mu_i^*\right)\varepsilon_{it}^* - \frac{1}{nT}\sum_{j=1}^{n}\sum_{s=1}^{T} u_{js-1}\varepsilon_{js}\right\} \\
&\quad - \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\left(y_{it-1}^* - \mu_i^*\right)\bar{\varepsilon}_i^* - \frac{1}{nT}\sum_{j=1}^{n}\sum_{s=1}^{T} u_{js-1}\bar{\varepsilon}_j\right\} \equiv B_{nT1}^* - B_{nT2}^*,
\end{aligned}$$

Therefore, it suffices to show that (b1) $B_{nT1}^* \to^{d^*} N(0, B)$ and (b2) $B_{nT2}^* \to^{P^*} 0$ in probability. For (b1), we verify the conditions of Theorem .1.2 with $B_{nT1}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{iT}^*$ and

$$Z_{iT}^* \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{it}^* \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(y_{it-1}^* - \mu_i^*\right)\varepsilon_{it}^* - \frac{1}{n\sqrt{T}}\sum_{j=1}^{n}\sum_{s=1}^{T} u_{js-1}\varepsilon_{js} \equiv q_{iT}^* - E^*\left(q_{iT}^*\right).$$

Notice that $\{Z_{iT}^*\}$ are independent across i for all T with $E^*\left(Z_{iT}^*\right) = 0$ and

$\Omega_{iT}^* \equiv E^*\left(Z_{iT}^{*2}\right) = E^*\left(q_{iT}^*\right)^2 - \left(E^*\left(q_{iT}^*\right)\right)^2$, where

$$\Omega^* \equiv \frac{1}{n}\sum_{i=1}^n \Omega_{iT}^* = \frac{1}{n}\sum_{i=1}^n E^*\left(q_{iT}^*\right)^2 - \frac{1}{n}\sum_{i=1}^n \left(E^*\left(q_{iT}^*\right)\right)^2 \equiv \Omega_1^* + \Omega_2^*.$$

By Lemma .1.4 (i),

$$\Omega_2^* = \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{n\sqrt{T}}\sum_{j=1}^n\sum_{s=1}^T u_{js-1}\varepsilon_{js}\right)^2 = \frac{1}{n}\left(\frac{1}{\sqrt{nT}}\sum_{j=1}^n\sum_{s=1}^T u_{js-1}\varepsilon_{js}\right)^2 = O_P\left(\frac{1}{n}\right).$$

Moreover,

$$\begin{aligned}\Omega_1^* &= \frac{1}{n}\sum_{i=1}^n E^*\left(q_{iT}^*\right)^2 = \frac{1}{n}\sum_{i=1}^n\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T u_{it-1}\varepsilon_{it}\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u_{it-1}^2\varepsilon_{it}^2 + \frac{1}{nT}\sum_{i=1}^n\sum_{t\neq s}^T u_{it-1}\varepsilon_{it}u_{is-1}\varepsilon_{is},\end{aligned}$$

where the first term converges to B in probability and the second term is an $o_P(1)$ given Assumption A1(vii) in particular. Thus, $\Omega^* \to^P B$. The result follows by showing that $E^*\left(q_{iT}^{*2+\delta}\right) = O(1)$ for $\delta = 2$. To prove (b2), we proceed similarly but verify the conditions of Theorem .1.1 instead. We omit the details to conserve space.

.1.3 Proofs of results in Section 1.4

Proof of Theorem 1.4.1. The proof follows from Theorem 1.3.1 and the fact that $\hat{\theta}_{rd}^* \to^{P^*} \theta_0$ in probability.

Proof of Theorem 1.4.2. The proof follows from Theorem 1.3.4 and the fact that $\hat{\theta}_{pb}^* \to^{P^*} \theta_0$ in probability.

Proof of Lemma 1.4.1. From the proof of Theorem 1.3.4, $\hat{A}_{pb}^* \to^{P^*} A$. Hence, it suffices to show that $\hat{B}_{pb}^* \to^{P^*} B$, in probability. We can write $\tilde{\varepsilon}_{it}^* - \bar{\tilde{\varepsilon}}_i^* = \hat{\varepsilon}_{it}^* - \bar{\hat{\varepsilon}}_i^* - \left(\hat{\hat{\theta}}_{pb}^* - \hat{\hat{\theta}}\right)\left(y_{it-1}^* - \bar{y}_{i-}^*\right)$, where $\tilde{\varepsilon}_{it}^* = y_{it}^* - \hat{\alpha}_i^* - \hat{\hat{\theta}}_{pb}^* y_{it-1}^*$ with $\hat{\alpha}_i^* = \bar{y}_i^* - \hat{\hat{\theta}}_{pb}^*\bar{y}_{i-}^*$ and $\hat{\varepsilon}_{it}^* = y_{it}^* - \check{\alpha}_i - \hat{\hat{\theta}} y_{it-1}^*$ with $\check{\alpha}_i = \hat{\alpha}_{I_i}$. As in the proof of Lemma 1.3.1, we can write

$$\hat{B}^*_{pb} = \hat{B}^*_1 + \hat{B}^*_2 + \hat{B}^*_3, \text{ with}$$

$\hat{B}^*_1 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i\right)^2$, $\hat{B}^*_2 = -2\left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right) \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\hat{\varepsilon}^*_{it} -\right.$
and $\hat{B}^*_3 = \left(\hat{\hat{\theta}}^*_{pb} - \hat{\hat{\theta}}\right)^2 \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$.
Given the pairs bootstrap DGP, one can show that $\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i\right)$
and
$\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$ are $O_{P^*}(1)$ terms, and therefore, $\hat{B}^*_2$ and $\hat{B}^*_3$ are $o_{P^*}(1)$ terms. For $\hat{B}^*_1$, we also define $\hat{\varepsilon}^*_{it} - \bar{\hat{\varepsilon}}^*_i = \varepsilon^*_{it} - \bar{\varepsilon}^*_i - \left(\hat{\hat{\theta}} - \theta_0\right)\left(y^*_{it-1} - \bar{y}^*_{i-}\right)$, where $\varepsilon^*_{it} = y^*_{it} - \alpha^*_i - \theta_0 y^*_{it-1}$ with $\tilde{\alpha}^*_i = \alpha_{I_i}$. This implies that

$$\hat{B}^*_1 = \chi^*_1 + \chi^*_2 + \chi^*_3, \text{ with}$$

$\chi^*_1 = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2$, $\chi^*_2 = -2\left(\hat{\hat{\theta}} - \theta_0\right) \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\varepsilon^*_{it} - \bar{\varepsilon}\right.$
and $\chi^*_3 = \left(\hat{\hat{\theta}} - \theta_0\right)^2 \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$. As before, one can show that given the bootstrap DGP of the pairwise bootstrap, $\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^3 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)$ and $\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^4$ are $O_{P^*}(1)$ and therefore, $\hat{\chi}^*_2$ and $\hat{\chi}^*_3$ are $o_{P^*}(1)$. Let us turn to $\hat{\chi}^*_1$. Since we have resampled only in the cross section,

$$\begin{aligned} E^*\left|\hat{\chi}^*_1\right| &= \frac{1}{n}\sum_{i=1}^n E^*\left\{\frac{1}{T}\sum_{t=1}^T \left(y^*_{it-1} - \bar{y}^*_{i-}\right)^2 \left(\varepsilon^*_{it} - \bar{\varepsilon}^*_i\right)^2\right\} = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y_{it-1} - \bar{y}_{i-}\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \\ &= \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y_{it-1} - \mu_i\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 + \frac{2}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(y_{it-1} - \mu_i\right)\left(\mu_i - \bar{y}_{i-}\right)\left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \\ &\quad + \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left(\mu_i - \bar{y}_{i-}\right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)^2 \equiv B_1 + B_2 + B_3, \end{aligned}$$

where $\mu_i = E(y_{it}) = \frac{\alpha_i}{1-\theta_0}$. By Cauchy-Schwartz inequality,

$$|B_2| \leq \left(\frac{2}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\mu_i - \bar{y}_{i-} \right) 2 \right)^{1/2} \left(\frac{2}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \mu_i \right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i \right)^4 \right)^{1/2} \to^P 0$$

since $\frac{2}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \mu_i \right)^2 \left(\varepsilon_{it} - \bar{\varepsilon}_i \right)^4 = O_P(1)$ and $\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(\mu_i - \bar{y}_{i-} \right) 2 = \frac{1}{n} \sum_{i=1}^{n} \bar{u}_{i-}^2 \to 0$ by Lemma .1.3. One can also show that $B_3 = o_P(1)$. Finally,

$$B_1 = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \mu_i \right)^2 \varepsilon_{it}^2 - \frac{2}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \mu_i \right)^2 \varepsilon_{it} \bar{\varepsilon}_i + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it-1} - \mu_i \right)^2 \bar{\varepsilon}_i^2$$

where the first term obviously converges in probability to B while the remaining terms converge to 0 by making use of the Cauchy-Schwartz inequality.

.2 Proof of main results in Chapter 2

.2.1 Proofs of results in Section 2.2

Proof of Theorem 2.2.1. Under Assumption A, Hahn and Kuersteiner (2011b) have shown that

$$\sqrt{nT} \left(\hat{\theta} - \theta_0 \right) = \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left(\frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} U_{it} \right)$$
$$-\sqrt{\frac{n}{T}} \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_i \right)^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial \gamma_i}\right]} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(U_{it}^{\gamma_i} - \frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial \gamma_i}\right]} v_{it} \right) \right] \right\} + o_P(1).$$

The half-panel jackknife estimator is given by $\hat{\theta}_{1/2} = 2\hat{\theta} - \frac{1}{2}\left(\hat{\theta}_1 + \hat{\theta}_2\right)$. If we assume for the mean time that $T_1 \equiv T/2$ is an integer, then by similar arguments, one can show that

$$\sqrt{nT_1}\left(\hat{\theta}_1-\theta_0\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=1}^{T_1}U_{it}\right)$$
$$-\sqrt{\frac{n}{T_1}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right]\right\}+o_P(1)$$

and

$$\sqrt{nT_1}\left(\hat{\theta}_2-\theta_0\right)=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T}U_{it}\right)-\sqrt{\frac{n}{T_1}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}$$
$$\times\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{\sqrt{T_1}}\sum_{t=T_1+1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T_1}}\sum_{t=T_1+1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right]\right\}+o_P(1).$$

By putting together these results, we obtain

$$\sqrt{nT}\left(\hat{\theta}_{1/2}-\theta_0\right)=\sqrt{nT}\left(2\hat{\theta}-\frac{1}{2}\left(\hat{\theta}_1+\hat{\theta}_2\right)-\theta_0\right)$$
$$=2\sqrt{nT}\left(\hat{\theta}-\theta_0\right)-\frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}_1-\theta_0\right)-\frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}_2-\theta_0\right)$$
$$=\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\right)-2\sqrt{\frac{n}{T}}\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}Z_{iT}+o_P(1),$$

where $Z_{iT}=Z_{1iT}+Z_{2iT}$, $T_1=T/2$ and

$$Z_{1iT}=\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_1}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=T_1+1}^{T}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right],$$
$$Z_{2iT}=\left[\frac{1}{\sqrt{T}}\sum_{t=T_1+1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_1}\left(U_{it}^{\gamma_i}-\frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2E\left[\frac{\partial v_{it}}{\partial\gamma_i}\right]}v_{it}\right)\right].$$

From Hahn and Kuersteiner (2011b), we have

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{I}_i\right)^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}U_{it}\rightarrow^{d}N\left(0,\mathcal{I}^{-1}\Omega\left(\mathcal{I}^{'}\right)^{-1}\right).$$

Thus, the half-panel jackknife estimator is asymptotically unbiased if $B_1 = \frac{1}{n}\sum_{i=1}^{n} Z_{1iT} = o_P(1)$ and $B_2 = \frac{1}{n}\sum_{i=1}^{n} Z_{2iT} = o_P(1)$. We are only going to give a proof of $B_1 = o_P(1)$, the proof of $B_2 = o_P(1)$ follows similarly. Note that the Z_{1iT} are independent across i and

$$E\left[Z_{1iT}\right] = \frac{\Sigma_{1iT}^{vU}}{E\left[\frac{\partial v_i}{\partial \gamma_i}\right]} - \frac{E\left(U_{it\gamma_i\gamma_i}\right)}{2\left(E\left[\frac{\partial v_i}{\partial \gamma_i}\right]\right)^2}\Sigma_{1iT}^{vv},$$

where $\Sigma_{1iT}^{vU} = \frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T} E\left[v_{it}U_{is}^{\gamma_i}\right]$ and $\Sigma_{1iT}^{vv} = \frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T} E\left[v_{it}v_{is}\right]$. Under Assumption A (3) and (4), we can apply Lemma 2.1 of Davydov (1968) with $p = q = 3$:

$$\begin{aligned}
\left\|\Sigma_{1iT}^{vU}\right\| &\leq \frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T}\left\|E\left[v_{it}U_{is}^{\gamma_i}\right]\right\| \\
&\leq \frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T} 12\left(E|v_{it}|^3\right)^{1/3}\left(E\|U_{is}^{\gamma_i}\|^3\right)^{1/3}\alpha_i\left(s-t\right)^{1-1/3-1/3} \\
&\leq 12\frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T}\left(E|M\left(x_{it}\right)|^3\right)^{1/3}\left(E|M\left(x_{it}\right)|^3\right)^{1/3}a^{(1/3)(s-t)} \\
&\leq \Delta\frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T} a^{1/3(s-t)} = \Delta a^{(1/3)}\left(\frac{1-a^{T_1/3}}{1-a^{1/3}}\right)^2\frac{1}{T} = O(\frac{1}{T}),
\end{aligned}$$

where Δ is a generic constant which follows from Assumption A (4). Therefore, $\Sigma_{1iT}^{vU} = o(1)$. Let us define $Z_{1iT,1} = \frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_i}{\partial \gamma_i}\right]}U_{is}^{\gamma_i}$. To conclude that

$$B_{1,1} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T}\frac{v_{it}}{E\left[\frac{\partial v_i}{\partial \gamma_i}\right]}U_{is}^{\gamma_i} = \frac{1}{n}\sum_{i=1}^{n}Z_{1iT,1} = o_P(1),$$

it suffices to show by Lemma 1 in Hansen (2007) that $E\left\|Z_{1iT,1}\right\|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T. We set $\delta = 1$. Then,

$$E\left\|Z_{1iT,1}\right\|^{2} = \frac{1}{T^2}\sum_{t,s=1}^{T_1}\sum_{p,q=T_1+1}^{T}\frac{E\left[v_{it}v_{ip}U_{is}^{\gamma_i'}U_{iq}^{\gamma_i}\right]}{\left(E\left[\frac{\partial v_i}{\partial \gamma_i}\right]\right)^2} = O(1),$$

by combining Lemma 1 of Andrews (1991) and Lemma 2.1 of Davydov (1968) [5]. Similarly, one can show that

$$B_{1,2} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T_1}\sum_{s=T_1+1}^{T}\frac{E\left(U_{is}^{\gamma_i\gamma_i}\right)}{2\left(E\left[\frac{\partial v_i}{\partial \gamma_i}\right]\right)^2}v_{it}v_{is} \equiv \frac{1}{n}\sum_{i=1}^{n}Z_{1iT,2} = o_P(1).$$

This complete the proof of $\frac{1}{n}\sum_{i=1}^{n}Z_{1iT} = o_P(1)$ and therefore, the half-panel jackknife estimator is asymptotically unbiased.

.2.2 Proofs of results in Section 2.3

In the bootstrap world, we define

$$\begin{aligned} u_{it}^{*}(\theta,\gamma_i) &= \frac{\partial\psi\left(x_{it}^{*};\theta,\gamma_i\right)}{\partial\theta}, \\ v_{it}^{*}(\theta,\gamma_i) &= \frac{\partial\psi\left(x_{it}^{*};\theta,\gamma_i\right)}{\partial\gamma_i}. \end{aligned}$$

[5] Arguments are similar to those of the proof of Theorem 1 of Hahn and Kuersteiner (2011b).

Therefore, the bootstrap counterpart of the maximization estimator is given by

$$\left(\hat{\theta}^{*},\hat{\gamma}_{1}^{*},\ldots,\hat{\gamma}_{n}^{*}\right)=\operatorname{argmax}_{\theta,\gamma_{1},\ldots,\gamma_{n}}\sum_{i=1}^{n}\sum_{t=1}^{T}\psi\left(x_{it}^{*};\theta,\gamma_{i}\right)$$

where $\{x_{it}^{*}, i=1,\ldots,n\ t=1,\ldots,T\}$ are data obtained by the pairs bootstrap procedure. Lemmas .2.1 and Lemma .2.2 are the analog of Lemmas A.4 and A.5 of Goncalves and White (2004). To state the following lemmas, we need to introduce some notation. For each $\theta\in\Theta$, let $Q_{nT}^{*}(\theta)\equiv\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}D^{\nu}\psi\left(x_{it}^{*};\theta,\hat{\gamma}_{i}^{*}(\theta)\right)$ be a pairs bootstrap resample of $Q_{nT}(\theta)\equiv\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}D^{\nu}\psi\left(x_{it};\theta,\hat{\gamma}_{i}(\theta)\right)$.

Lemma .2.1. ***(Bootstrap Pointwise WLLN).*** *Under Assumption A, for any $\eta>0$, $\delta>0$ and for each $\theta\in\Theta$,*

$$\lim_{n\to\infty}P\left[P^{*}\left(\left|Q_{nT}^{*}(\theta)-Q_{nT}(\theta)\right|>\eta\right)>\delta\right]=0.$$

Lemma .2.2. ***(Bootstrap Uniform WLLN).*** *Under Assumption A, for any $\eta>0$ and $\delta>0$,*

$$\lim_{n\to\infty}P\left[P^{*}\left(\sup_{\theta\in\Theta}\left|Q_{nT}^{*}(\theta)-Q_{nT}(\theta)\right|>\eta\right)>\delta\right]=0.$$

Our next lemma establishes the consistency of $\hat{\theta}^{*}$.

Lemma .2.3. ***(Consistency of $\hat{\theta}^{*}$).*** *Under Assumption A,*

$$\hat{\theta}^{*}-\hat{\theta}\to^{P^{*}}0.$$

Before stating our next lemma, let

$$\mathcal{I}_{nT}^{*}\left(\tilde{\theta}^{*}\right)=-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)-\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}v_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)\right\},$$

where $\tilde{\theta}^{*}$ lies between $\hat{\theta}^{*}$ and $\hat{\theta}$. The next lemma establishes the consistency of the Jacobian.

Lemma .2.4. ***(Consistency of the Jacobian).*** *Under Assumption A and for every* $\tilde{\theta}^* \to^{P^*} \theta_0$*,*

$$\mathcal{I}^*_{nT}\left(\tilde{\theta}^*\right) - \mathcal{I} \to^{P^*} 0,$$

where $\mathcal{I} \equiv \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathcal{I}_i$.

For the next Lemma, we define $D^*_{nT} \equiv \frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right)$.

Lemma .2.5. ***(Asymptotic Normality).*** *Under Assumption A,*

$$D^*_{nT} \to^{d^*} N\left(0, \Omega\right), \text{ in probability,}$$

with Ω *defined as in Theorem 2.2.1.*

Proof of Theorem 2.3.1. Recall that the pairs bootstrap resamples only in the cross sectional dimension. More specifically, we generate $x^*_i \sim$ i.i.d. $\{x_i : i = 1, \ldots, n\}$, where $x^*_i = (x^*_{i1}, \ldots, x^*_{iT})$ and $x_i = (x_{i1}, \ldots, x_{iT})$; i.e. letting $I_1, \ldots, I_n$ be i.i.d. Uniform on $\{1, \ldots, n\}$, we have $x^*_{it} = x_{I_i t}$. Thus, $\gamma^*_{i0} \equiv \hat{\gamma}_{I_i}$ is the bootstrap counterpart of γ_{i0}. Let $\hat{\gamma}^*_i(\theta) \equiv \text{argmax}_a \sum_{t=1}^T \psi\left(x^*_{it}; \theta, a\right)$. Notice that given the bootstrap DGP, $\hat{\gamma}^*_i(\theta) = \hat{\gamma}_{I_i}(\theta)$ for all θ where $\hat{\gamma}_i(\theta) \equiv \text{argmax}_a \sum_{t=1}^T \psi\left(x_{it}; \theta, a\right)$. The following first order condition (FOC) holds:

$$\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T u^*_{it}\left(\hat{\theta}^*, \hat{\gamma}^*_i\left(\hat{\theta}^*\right)\right) = 0.$$

Expanding this expression around the true parameter value $\hat{\theta}$ (in the bootstrap world) yields

$$0 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}^{*}\left(\hat{\theta}, \hat{\gamma}_i^{*}\left(\hat{\theta}\right)\right)$$
$$+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right) + u_{it\gamma_i}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)\frac{\partial\hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)}{\partial\theta}\right\}\left(\hat{\theta}^{*} - \hat{\theta}\right)$$

where $\tilde{\theta}^{*}$ lies between $\hat{\theta}^{*}$ and $\hat{\theta}$. For the estimators of the individual effects, the first order condition is $\frac{1}{T}\sum_{t=1}^{T} v_{it}^{*}\left(\theta, \hat{\gamma}_i^{*}\left(\theta\right)\right) = 0$. Differentiating this expression with respect to θ yields:

$$\frac{\partial\hat{\gamma}_i^{*}\left(\theta\right)}{\partial\theta} = -\frac{\frac{1}{T}\sum_{t=1}^{T} v_{it\theta}^{*}\left(\theta, \hat{\gamma}_i^{*}\left(\theta\right)\right)}{\frac{1}{T}\sum_{t=1}^{T} v_{it\gamma_i}^{*}\left(\theta, \hat{\gamma}_i^{*}\left(\theta\right)\right)}.$$

Therefore, we can write

$$\begin{aligned}\hat{\theta}^{*} - \hat{\theta} &= -\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right) - u_{it\gamma_i}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)\frac{\frac{1}{T}\sum_{t=1}^{T} v_{it\theta}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T} v_{it\gamma_i}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)}\right\}\right]^{-1} \\ &\quad \times\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}^{*}\left(\hat{\theta}, \hat{\gamma}_i^{*}\left(\hat{\theta}\right)\right),\end{aligned}$$

or alternatively,

$$\begin{aligned}\sqrt{nT}\left(\hat{\theta}^{*} - \hat{\theta}\right) &= -\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right) - \frac{\frac{1}{T}\sum_{t=1}^{T} u_{it\gamma_i}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T} v_{it\gamma_i}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)} v_{it\theta}^{*}\left(\tilde{\theta}^{*}, \hat{\gamma}_i^{*}\left(\tilde{\theta}^{*}\right)\right)\right\}\right]^{-1} \\ &\quad \times\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} u_{it}^{*}\left(\hat{\theta}, \gamma_{i0}^{*}\right),\end{aligned}$$

since by definition $\hat{\gamma}_i^{*}\left(\hat{\theta}\right) = \hat{\gamma}_{I_i}\left(\hat{\theta}\right) \equiv \gamma_{i0}^{*}$. The result follows from Lemma

.2.4 and lemma .2.5 since they establish that

$$\mathcal{I}_{nT}^{*}\left(\tilde{\theta}^{*}\right)=-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{u_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)-\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_{i}}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)}v_{it\theta}^{*}\left(\tilde{\theta}^{*},\hat{\gamma}_{i}^{*}\left(\tilde{\theta}^{*}\right)\right)\right\}\rightarrow^{P^{*}}\mathcal{I}$$

and

$$D_{nT}^{*}=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^{*}\left(\hat{\theta},\gamma_{i0}^{*}\right)\rightarrow^{d^{*}}N\left(0,\Omega\right),\text{ in probability.}$$

Proof of Lemma .2.1 Define $Z_{iT}^{*}\left(\theta\right)\equiv\frac{1}{T}\sum_{t=1}^{T}D^{\nu}\psi\left(x_{it}^{*};\theta,\hat{\gamma}_{i}^{*}\left(\theta\right)\right)$. Then $Q_{nT}^{*}\left(\theta\right)\equiv\frac{1}{n}\sum_{i=1}^{n}Z_{iT}^{*}$. Since $\hat{\gamma}_{i}^{*}\left(\theta\right)$ and $\hat{\gamma}_{i}\left(\theta\right)$ depend only on observations of individuals I_i and i respectively, it follows that $E^{*}\left(Q_{nT}^{*}\left(\theta\right)\right)=E^{*}\left(Z_{iT}^{*}\left(\theta\right)\right)=Q_{nT}\left(\theta\right)$ for all $\theta\in\Theta$. Conditionally on the data, we have by Tchebychev's inequality

$$\begin{aligned}
&P^{*}\left(\left|Q_{nT}^{*}\left(\theta\right)-Q_{nT}\left(\theta\right)\right|\geq\eta\right)\leq\frac{1}{\eta^{2}}Var^{*}\left(Q_{nT}^{*}\left(\theta\right)-Q_{nT}\left(\theta\right)\right)\\
&=\frac{1}{\eta^{2}}Var^{*}\left(\frac{1}{n}\sum_{i=1}^{n}Z_{iT}^{*}\left(\theta\right)-E^{*}\left(Z_{iT}^{*}\left(\theta\right)\right)\right)=\frac{1}{\eta^{2}}\frac{1}{n^{2}}\sum_{i=1}^{n}Var^{*}\left(Z_{iT}^{*}\left(\theta\right)-Q_{nT}\left(\theta\right)\right)\\
&=\frac{1}{\eta^{2}}\frac{1}{n^{2}}\sum_{i=1}^{n}\left\{E^{*}\left(Z_{iT}^{*}\left(\theta\right)Z_{iT}^{*}\left(\theta\right)'\right)-Q_{nT}\left(\theta\right)Q_{nT}\left(\theta\right)'\right\}\\
&=\frac{1}{\eta^{2}}\frac{1}{n^{2}}\sum_{i=1}^{n}\left\{\frac{1}{T^{2}}\sum_{t,s=1}^{T}E^{*}\left(D^{\nu}\psi\left(x_{it}^{*};\theta,\hat{\gamma}_{i}^{*}\left(\theta\right)\right)D^{\nu}\psi\left(x_{is}^{*};\theta,\hat{\gamma}_{i}^{*}\left(\theta\right)\right)'\right)\right\}-\frac{1}{\eta^{2}}\frac{1}{n}Q_{nT}\left(\theta\right)Q_{nT}\left(\theta\right)'\\
&=\frac{1}{\eta^{2}}\frac{1}{n^{2}T^{2}}\sum_{i=1}^{n}\sum_{t,s=1}^{T}D^{\nu}\psi\left(x_{it};\theta,\hat{\gamma}_{i}\left(\theta\right)\right)D^{\nu}\psi\left(x_{is};\theta,\hat{\gamma}_{i}\left(\theta\right)\right)'-\frac{1}{\eta^{2}}\frac{1}{n}Q_{nT}\left(\theta\right)Q_{nT}\left(\theta\right)'=O_{P}\left(\frac{1}{n}\right),
\end{aligned}$$

where the last equality follows from Assumption A (4), delivering the desired result.

Proof of Lemma .2.2 Given $\eta>0$, divide Θ into subsets $\Theta_1,\Theta_2,\ldots,\Theta_{M(\varepsilon)}$ such that $\|\theta_1-\theta_2\|<\varepsilon$ whenever θ_1 and θ_2 are in the same subsets. Let θ_i

be a point in Θ_i for each $i = 1, \ldots, M(\eta)$. Then

$$P^* \left(\sup_{\theta \in \Theta} |Q_{nT}^*(\theta) - Q_{nT}(\theta)| \geq \eta \right) = P^* \left(\max_j \sup_{\theta \in \Theta_j} |Q_{nT}^*(\theta) - Q_{nT}(\theta)| \geq \eta \right)$$
$$\leq \sum_{j=1}^{M(\eta)} P^* \left(\sup_{\theta \in \Theta_j} |Q_{nT}^*(\theta) - Q_{nT}(\theta)| \geq \eta \right).$$

For $\theta \in \Theta_j$,

$$|Q_{nT}^*(\theta) - Q_{nT}(\theta)| \leq |Q_{nT}^*(\theta) - Q_{nT}^*(\theta_j)| + |Q_{nT}^*(\theta_j) - Q_{nT}(\theta_j)| + |Q_{nT}(\theta_j) - Q_{nT}(\theta)|$$
$$\leq |Q_{nT}^*(\theta_j) - Q_{nT}(\theta_j)| + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}^*) (\varepsilon + \max_i |\hat{\gamma}_i^*(\theta) - \hat{\gamma}_i^*(\theta_j)|)$$
$$+ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}) (\varepsilon + \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|)$$
$$\leq |Q_{nT}^*(\theta_j) - Q_{nT}(\theta_j)| + \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}^*) (\varepsilon + \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|)$$
$$+ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}) (\varepsilon + \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|)$$
$$\leq |Q_{nT}^*(\theta_j) - Q_{nT}(\theta_j)| + \frac{2}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}) (\varepsilon + \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|)$$
$$+ \left[\frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}^*) - \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} M(x_{it}) \right] (\varepsilon + \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|),$$

where the third inequality uses $\max_i |\hat{\gamma}_i^*(\theta) - \hat{\gamma}_i^*(\theta_j)| \leq \max_i |\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)|$ given that $\hat{\gamma}_i^*(\theta) = \hat{\gamma}_{I_i}(\theta)$ for all θ. Since

$$\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta) = \frac{\partial \hat{\gamma}_i\left(\tilde{\theta}\right)}{\partial \theta} (\theta_j - \theta) = -\frac{\frac{1}{T} \sum_{t=1}^{T} v_{it\theta}\left(\tilde{\theta}, \hat{\gamma}_i\left(\tilde{\theta}\right)\right)}{\frac{1}{T} \sum_{t=1}^{T} v_{it\gamma_i}\left(\tilde{\theta}, \hat{\gamma}_i\left(\tilde{\theta}\right)\right)} (\theta_j - \theta),$$

where $\tilde{\theta}$ lies between θ_j and θ, we have

$$\max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right| \le \varepsilon \max_i \left| \frac{\frac{1}{T}\sum_{t=1}^{T} v_{it\theta}\left(\tilde{\theta}, \hat{\gamma}_i\left(\tilde{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T} v_{it\gamma_i}\left(\tilde{\theta}, \hat{\gamma}_i\left(\tilde{\theta}\right)\right)} \right| = O_P(\varepsilon),$$

under Assumption A (4). Choose $\varepsilon > 0$ such that

$$\left| \frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x_{it})\left(\varepsilon + \max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right|\right) \right| < \frac{\eta}{3} \text{ w.p.a. 1.}$$

Notice that this is always possible since we have shown that $\max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right| = O_P(\varepsilon)$. Then

$$P^*\left(\sup_{\theta\in\Theta_j} \left|Q^*_{nT}(\theta) - Q_{nT}(\theta)\right| \ge \eta\right) \le P^*\left(\left|Q^*_{nT}(\theta_j) - Q_{nT}(\theta_j)\right| \ge \frac{\eta}{3}\right)$$
$$+P^*\left(\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x^*_{it}) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x_{it})\right| \left(\varepsilon + \max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right|\right) \ge \frac{\eta}{3}\right)$$
$$+ o_P(1).$$

We know from Lemma .2.1 that $P^*\left(\left|Q^*_{nT}(\theta_j) - Q_{nT}(\theta_j)\right| \ge \frac{\eta}{3}\right) = o_P(1)$. On the other hand, to show that

$$P^*\left(\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x^*_{it}) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x_{it})\right| \left(\varepsilon + \max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right|\right) \ge \frac{\eta}{3}\right) = o_P(1),$$

it suffices to check that $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x^*_{it}) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} M(x_{it}) = o_{P^*}(1)$ since $\max_i \left|\hat{\gamma}_i(\theta_j) - \hat{\gamma}_i(\theta)\right| = O_P(\varepsilon)$ and $\varepsilon > 0$ is fixed. However, this result follows from arguments similar to those in Lemma .2.1. This complete the proof of Lemma .2.2.

Proof of Lemma .2.3 Under Assumption A, $Q_{nT}(\theta)$ is uniquely maximized on Θ at $\hat{\theta}$ w.p.a. 1. Then by standard arguments for extremum estimators, Lemma .2.2 applied with $Q^*_{nT}(\theta) \equiv \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \psi\left(x^*_{it};\theta,\hat{\gamma}^*_i(\theta)\right)$ delivers the desired result.

Proof of Lemma .2.4 The Jacobian can be rewritten as

$$\begin{aligned}
\mathcal{I}^*_{nT}\left(\tilde{\theta}^*\right) &= -\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T \left\{ u^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right) - \frac{\frac{1}{T}\sum_{t=1}^T u^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)}{\frac{1}{T}\sum_{t=1}^T v^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)} v^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)\right\} \\
&\equiv -\mathcal{J}_1\left(\tilde{\theta}^*\right) + \mathcal{J}_2\left(\tilde{\theta}^*\right),
\end{aligned}$$

with the obvious definitions. For $\mathcal{J}_1\left(\tilde{\theta}^*\right)$, we can write

$$\mathcal{J}_1\left(\tilde{\theta}^*\right) = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right) - \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)$$
$$+\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right) - \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right) + \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right).$$

Lemma .2.1 shows that

$$\frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right) - \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right) = o_{P^*}(1).$$

Also, by Assumption A (4),

$$\begin{aligned}
&\left\| \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T u^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right) - u^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right) \right\| \\
&\quad \le \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T M\left(x^*_{it}\right)\left(\left\|\tilde{\theta}^* - \hat{\theta}\right\| + \max_i \left|\hat{\gamma}^*_i\left(\tilde{\theta}^*\right) - \hat{\gamma}^*_i\left(\hat{\theta}\right)\right|\right) = o_{P^*}(1)
\end{aligned}$$

since $\hat{\theta}^* - \hat{\theta} = o_{P^*}(1)$ and $\tilde{\theta}^*$ lies between $\hat{\theta}^*$ and $\hat{\theta}$. In the last statement, the less obvious result is $\max_i \left|\hat{\gamma}^*_i\left(\tilde{\theta}^*\right) - \hat{\gamma}^*_i\left(\hat{\theta}\right)\right| = o_{P^*}(1)$ but this follows by the mean value theorem since

$$\hat{\gamma}_i^*\left(\tilde{\theta}^*\right)-\hat{\gamma}_i^*\left(\hat{\theta}\right)=\frac{\partial\hat{\gamma}_i^*\left(\bar{\theta}^*\right)}{\partial\theta}\left(\tilde{\theta}^*-\hat{\theta}\right)=-\frac{\frac{1}{T}\sum_{t=1}^{T}v_{it\theta}^*\left(\bar{\theta}^*,\hat{\gamma}_i^*\left(\bar{\theta}^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}^*\left(\bar{\theta}^*,\hat{\gamma}_i^*\left(\bar{\theta}^*\right)\right)}\left(\tilde{\theta}^*-\hat{\theta}\right)$$

where $\bar{\theta}^*$ lies between $\tilde{\theta}^*$ and $\hat{\theta}$. The result follows given that $\frac{\frac{1}{T}\sum_{t=1}^{T}v_{it\theta}^*\left(\bar{\theta}^*,\hat{\gamma}_i^*\left(\bar{\theta}^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}^*\left(\bar{\theta}^*,\hat{\gamma}_i^*\left(\bar{\theta}^*\right)\right)}=O_{P^*}(1)$ uniformly on i thanks to Assumption A (4) and (6). Therefore, we have shown that

$$\mathcal{J}_1\left(\tilde{\theta}^*\right)=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)+o_{P^*}(1).$$

For $\mathcal{J}_2\left(\tilde{\theta}^*\right)$, we can apply the kind of arguments we have used for $\mathcal{J}_1\left(\tilde{\theta}^*\right)$ by replacing therein $u_{it\theta}^*\left(\tilde{\theta}^*,\hat{\gamma}_i^*\left(\tilde{\theta}^*\right)\right)$ by

$$\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_i}^*\left(\hat{\theta},\hat{\gamma}_i^*\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}^*\left(\hat{\theta},\hat{\gamma}_i^*\left(\hat{\theta}\right)\right)}v_{it\theta}^*\left(\tilde{\theta}^*,\hat{\gamma}_i^*\left(\tilde{\theta}^*\right)\right).$$

$$\mathcal{J}_2\left(\tilde{\theta}^*\right) = \underbrace{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left\{\frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)} - \frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}\right\}v^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)}_{\mathcal{J}^*_{21}}$$

$$+\underbrace{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}\left(v^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right) - v^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)\right)}_{\mathcal{J}^*_{22}}$$

$$+\underbrace{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}v^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right) - \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}v_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}_{\mathcal{J}^*_{23}}$$

$$+\underbrace{\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}v_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}_{\mathcal{J}_{24}}.$$

$\mathcal{J}^*_{21}$ and $\mathcal{J}^*_{22}$ are $o_{P^*}(1)$ terms since

$$\begin{aligned}\max_i &\left|\frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right)} - \frac{\frac{1}{T}\sum_{t=1}^{T}u^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v^*_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)}\right| \\ &= O_{P^*}\left(\left\|\tilde{\theta}^*-\hat{\theta}\right\| + \max_i\left|\hat{\gamma}^*_i\left(\tilde{\theta}^*\right) - \hat{\gamma}^*_i\left(\hat{\theta}\right)\right|\right) = o_{P^*}(1),\end{aligned}$$

and

$$\begin{aligned}\max_i &\left|\frac{1}{T}\sum_{t=1}^{T}v^*_{it\theta}\left(\tilde{\theta}^*,\hat{\gamma}^*_i\left(\tilde{\theta}^*\right)\right) - \frac{1}{T}\sum_{t=1}^{T}v^*_{it\theta}\left(\hat{\theta},\hat{\gamma}^*_i\left(\hat{\theta}\right)\right)\right| \\ &= O_{P^*}\left(\left\|\tilde{\theta}^*-\hat{\theta}\right\| + \max_i\left|\hat{\gamma}^*_i\left(\tilde{\theta}^*\right) - \hat{\gamma}^*_i\left(\hat{\theta}\right)\right|\right) = o_{P^*}(1),\end{aligned}$$

given Assumption A (4) and (6) and the fact that $\hat{\theta}^* - \hat{\theta} = o_{P^*}(1)$ and $\tilde{\theta}^*$ lies between $\hat{\theta}^*$ and $\hat{\theta}$. For $\mathcal{J}^*_{23}$, one can apply the same kind of arguments as in the proof of Lemma .2.1 to show that $\mathcal{J}^*_{23} = o_{P^*}(1)$. Thus, we have shown that

$$\mathcal{J}_2\left(\tilde{\theta}^*\right) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}v_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right) + o_{P^*}(1)$$

and therefore

$$\begin{aligned}\mathcal{I}^*_{nT}\left(\tilde{\theta}^*\right) &= -\mathcal{J}_1\left(\tilde{\theta}^*\right) + \mathcal{J}_2\left(\tilde{\theta}^*\right) \\ &= -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left[u_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right) - \frac{\frac{1}{T}\sum_{t=1}^{T}u_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}{\frac{1}{T}\sum_{t=1}^{T}v_{it\gamma_i}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)}v_{it\theta}\left(\hat{\theta},\hat{\gamma}_i\left(\hat{\theta}\right)\right)\right] + o_{P^*}(1) \\ &= -\frac{1}{n}\sum_{i=1}^{n}\hat{\mathcal{I}}_i + o_{P^*}(1) \to^{P^*} \mathcal{I},\end{aligned}$$

by Lemma 13 of Hahn and Kuersteiner (2011b).

Proof of Lemma .2.5 We define

$$D^*_{nT} \equiv \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u^*_{it}\left(\hat{\theta},\gamma^*_{i0}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}D^*_{iT},$$

where $D^*_{iT} \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u^*_{it}\left(\hat{\theta},\gamma^*_{i0}\right)$. We are going to check the conditions of Lemma 2 in Hansen (2007). Note that given our bootstrap DGP and recalling that $\gamma^*_{i0} = \hat{\gamma}_{I_i}\left(\hat{\theta}\right) = \hat{\gamma}_{I_i}$, the D^*_{it} are independent across i and

$$E^*\left(D^*_{iT}\right) = \frac{1}{n\sqrt{T}}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}\left(\hat{\theta},\hat{\gamma}_i\right) = 0$$

by the FOC. This last result explains why the pairs bootstrap is not able to mimic well the incidental parameter bias. Also,

$$\begin{aligned} E^*\left(D_{iT}^* D_{iT}^{*\prime}\right) &= \frac{1}{T}\sum_{t,s=1}^{T} E^*\left(u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right)u_{is}^*\left(\hat{\theta},\gamma_{i0}^*\right)'\right) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t,s=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)u_{is}\left(\hat{\theta},\hat{\gamma}_i\right)' \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)\right)'. \end{aligned}$$

Expanding $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right)$ around (θ_0,γ_{i0}) and using $\hat{\theta}-\theta_0 = O_P\left(\frac{1}{\sqrt{nT}}\right) = O_P\left(\frac{1}{T}\right)$[6] give

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it\theta}\left(\hat{\theta}-\theta_0\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it\gamma_i}\left(\hat{\gamma}_i-\gamma_{i0}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it} + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it\gamma_i}\left(\hat{\gamma}_i-\gamma_{i0}\right) + O_P\left(\frac{1}{\sqrt{T}}\right). \end{aligned}$$

Since $\hat{\gamma}_i-\gamma_{i0} = -\frac{\frac{1}{T}\sum_{t=1}^{T} v_{it}}{E\left(v_{it\gamma_i}\right)} + O_P\left(\frac{1}{T}\right)$[7], we can write

$$\begin{aligned} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}\left(\hat{\theta},\hat{\gamma}_i\right) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(u_{it} - \frac{\frac{1}{T}\sum_{t=1}^{T} u_{it\gamma_i}}{E\left(v_{it\gamma_i}\right)} v_{it}\right) + O_P\left(\frac{1}{\sqrt{T}}\right) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it} + O_P\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

since $\frac{1}{T}\sum_{t=1}^{T} u_{it\gamma_i} = E\left(u_{it\gamma_i}\right) + O_P\left(\frac{1}{\sqrt{T}}\right)$. Thus,

$$\begin{aligned} E^*\left(D_{iT}^* D_{iT}^{*\prime}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it} + O_P\left(\frac{1}{\sqrt{T}}\right)\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it} + O_P\left(\frac{1}{\sqrt{T}}\right)\right)' \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U_{it}\right)' + O_P\left(\frac{1}{\sqrt{T}}\right) \to^P \Omega \end{aligned}$$

[6] We are using the fact that $T = O(n)$.

[7] see Kim and Sun (2013), P.30.

under Hahn and Kuersteiner (2011b) assumptions. To conclude that $D^*_{nT} \to^{d^*} N(0,\Omega)$ it suffices then to show that $E^*\left(\|D^*_{iT}\|^4\right) = O_P(1)$, or alternatively that $E^*\left(\left|D^{*j}_{iT}\right|^4\right) = O_P(1)$, where D^{*j}_{iT} is the j^{th} element of D^*_{iT}, $j = 1, \ldots, p$. Similarly, for fixed j,

$$\begin{aligned} E^*\left(\left|D^{*j}_{iT}\right|^4\right) &= E^*\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u^{*j}_{it}\left(\hat{\theta}, \gamma^*_{i0}\right)\right|^4\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u^{j}_{it}\left(\hat{\theta}, \hat{\gamma}_i\right)\right)^4 \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U^{j}_{it} + O_P\left(\frac{1}{\sqrt{T}}\right)\right)^4 \\ &\leq C\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U^{j}_{it}\right)^4 + O_P\left(\frac{1}{T^2}\right), \end{aligned}$$

where C is a constant which do not depend on n and T. The result follows since $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} U^{j}_{it}\right)^4 = O_P(1)$ by arguments similar to those in the proof of Theorem 1 in Hahn and Kuersteiner (2011b).

.2.3 Proofs of results in Section 2.4

Proof of Theorem 2.4.1 We first assume that T is even and we define $T_1 = T/2$. By replacing $\hat{\theta}^*_{1/2}$ and $\hat{\theta}_{1/2}$ with (2.3) and (2.1) respectively, we have that

$$\sqrt{nT}\left(\hat{\theta}^*_{1/2} - \hat{\theta}_{1/2}\right) = 2\sqrt{nT}\left(\hat{\theta}^* - \hat{\theta}\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_1 - \hat{\theta}_1\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_2 - \hat{\theta}_2\right).$$

By Theorem 2.3.1, we know that the following expansions hold

$$\begin{aligned}\sqrt{nT_1}\left(\hat{\theta}_1^* - \hat{\theta}_1\right) &= \mathcal{I}_{1nT_1}^{*-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=1}^{T_1}u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right)\\ \sqrt{nT_1}\left(\hat{\theta}_2^* - \hat{\theta}_2\right) &= \mathcal{I}_{2nT_1}^{*-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T}u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right),\end{aligned}$$

where

$$\begin{aligned}\mathcal{I}_{1nT_1}^* &= -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T_1}\left\{u_{it\theta}^*\left(\tilde{\theta}_1^*,\hat{\gamma}_i^*\left(\tilde{\theta}_1^*\right)\right) - \frac{\frac{1}{T}\sum_{t=1}^{T_1}u_{it\gamma_i}^*\left(\tilde{\theta}_1^*,\hat{\gamma}_i^*\left(\tilde{\theta}_1^*\right)\right)}{\frac{1}{T}\sum_{t=1}^{T_1}v_{it\gamma_i}^*\left(\tilde{\theta}_1^*,\hat{\gamma}_i^*\left(\tilde{\theta}_1^*\right)\right)}v_{it\theta}^*\left(\tilde{\theta}_1^*,\hat{\gamma}_i^*\left(\tilde{\theta}_1^*\right)\right)\right\}\\ \mathcal{I}_{2nT_1}^* &= -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T}\left\{u_{it\theta}^*\left(\tilde{\theta}_2^*,\hat{\gamma}_i^*\left(\tilde{\theta}_2^*\right)\right) - \frac{\frac{1}{T}\sum_{t=T_1+1}^{T}u_{it\gamma_i}^*\left(\tilde{\theta}_2^*,\hat{\gamma}_i^*\left(\tilde{\theta}_2^*\right)\right)}{\frac{1}{T}\sum_{t=T_1+1}^{T}v_{it\gamma_i}^*\left(\tilde{\theta}_2^*,\hat{\gamma}_i^*\left(\tilde{\theta}_2^*\right)\right)}v_{it\theta}^*\left(\tilde{\theta}_2^*,\hat{\gamma}_i^*\left(\tilde{\theta}_2^*\right)\right)\right\},\end{aligned}$$

with $\tilde{\theta}_j^*$ lying between $\hat{\theta}_j^*$ and $\hat{\theta}_j$, $j = 1, 2$. Since from Lemma .2.4 $\mathcal{I}_{nT}^* \to^{P^*} \mathcal{I}$, $\mathcal{I}_{1nT_1}^* \to^{P^*} \mathcal{I}$ and $\mathcal{I}_{2nT_1}^* \to^{P^*} \mathcal{I}$, we can write

$$\begin{aligned}\sqrt{nT_1}\left(\hat{\theta}_1^* - \hat{\theta}_1\right) &= \mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=1}^{T_1}u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right) + o_{P^*}(1)\\ \sqrt{nT_1}\left(\hat{\theta}_2^* - \hat{\theta}_2\right) &= \mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T}u_{it}^*\left(\hat{\theta},\gamma_{i0}^*\right) + o_{P^*}(1).\end{aligned}$$

Thus, since $T_1 = T/2$

$$\begin{aligned}
\sqrt{nT}\left(\hat{\theta}^*_{1/2} - \hat{\theta}_{1/2}\right) &= 2\sqrt{nT}\left(\hat{\theta}^* - \hat{\theta}\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_1 - \hat{\theta}_1\right) - \frac{\sqrt{2}}{2}\sqrt{nT_1}\left(\hat{\theta}^*_2 - \hat{\theta}_2\right) \\
&= 2\mathcal{I}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) - \frac{\sqrt{2}}{2}\mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=1}^{T_1}u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) \\
&\quad - \frac{\sqrt{2}}{2}\mathcal{I}^{-1}\frac{1}{\sqrt{nT_1}}\sum_{i=1}^{n}\sum_{t=T_1+1}^{T}u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) + o_{P^*}(1) \\
&= \mathcal{I}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}u^*_{it}\left(\hat{\theta}, \gamma^*_{i0}\right) + o_{P^*}(1) \to^{d^*} N\left(0, \mathcal{I}^{-1}\Omega\mathcal{I}^{-1}\right),
\end{aligned}$$

where the limit in distribution follows from the proof of Theorem 2.3.1.

.3 Proof of main results in Chapter 3

.3.1 Proofs of results for the Standard Bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix A. Let $\hat{\beta}^* = \hat{\beta}^*_{std}$ throughout Appendix A. Let C denote a generic positive constant that may be different in different uses. Also, P^* denotes the probability measure induced by the standard residual based bootstrap procedure and E^* denotes the expectation under P^*.

Lemma .3.1. *Suppose that Assumptions 1-2 hold, then (a)* $E^*\left(\epsilon_i^{*8}\right)$ *and (b)* $E^*\left(\|V_i^*\|^8\right)$ *are bounded in probability.*

Lemma .3.2. *Suppose that Assumptions 1-2 hold, then the following statements are true:*

(a) $V^{*\prime}P_Z\epsilon^*/l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, *(b)* $V^{*\prime}P_Z V^*/l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right)$, *(c)* $\epsilon^{*\prime}P_Z\epsilon^*/l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$, *in probability; in Case (I), (d)* $\hat{\Pi}'Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, *(e)* $\hat{\Pi}'Z'\epsilon^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, *in probability; in Case (II),*

(d') $\hat{\Pi}'Z'V^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, (e') $\hat{\Pi}'Z'\epsilon^*/l = O_{P^*}\left(1/\sqrt{l}\right)$, *in probability; where* $\sigma^b_{V\epsilon} \equiv E^*\left(V_i^*\epsilon_i^*\right)$, $\Sigma^b_{VV} \equiv E^*\left(V_i^*V_i^{*'}\right)$ *and* $\sigma^b_{\epsilon\epsilon} \equiv E^*\left(\epsilon_i^{*2}\right)$.

To proceed, let $A^* = diag(a_1^*, ..., a_n^*)$ where $a_i^* = \epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}, i = 1, .., n$.

Lemma .3.3. *Suppose that Assumptions 1-2 hold, then the following statements are true:* (a) $\tilde{V}^{*'}A^*\tilde{V}^*/n = E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (b) $\tilde{V}^{*'}P_Z A^*\tilde{V}^*/n = \lambda_n E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (c) $\tilde{V}^{*'}D_Z A^* D_Z\tilde{V}^*/n = \lambda_n\phi_n E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$; (d) $\tilde{V}^{*'}P_Z A^* P_Z\tilde{V}^*/n = \lambda_n\phi_n E^*\left(a_i^*\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(1/\sqrt{n}\right)$, *in probability.*

Lemma .3.4. *Suppose that Assumptions 1-2 hold, then both in Case (I) and Case (II),* $\hat{\beta}^* - \hat{\beta} = o_{P^*}(1)$, *in probability.*

Let $\hat{\lambda}^*(\hat{\beta}^*) = \frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)P_Z\hat{\epsilon}^*(\hat{\beta}^*)}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}$, $\lambda^* = \frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}\epsilon^*}$ where $\hat{\epsilon}^*(\hat{\beta}^*) = y^* - X^*\hat{\beta}^*$ and $\{y^*, X^*\}$ denotes the pseudo-sample generated by the standard bootstrap.

Lemma .3.5. *Suppose that Assumptions 1-2 hold, then* $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$.

Lemma .3.6. *If* $\hat{\lambda}^*(\hat{\beta}^*) = \lambda^* + O_{P^*}\left(\delta_n^\lambda\right)$ *for* $\delta_n^\lambda \to 0$, *and* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n^\beta \to 0$, *then for Case (I),* (a) $\frac{1}{r_n}(X^{*'}P_Z X^* - \hat{\lambda}^*(\hat{\beta}^*)X^{*'}X^*) = \bar{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\lambda n/r_n\right)$, (b) $\frac{1}{r_n}(X^{*'}P_Z\hat{\epsilon}^* - \hat{\lambda}^*(\hat{\beta}^*)X^{*'}\hat{\epsilon}^*) = O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\beta + \delta_n^\lambda n/r_n\right)$, *where* $\bar{H}_{I,n} = (1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/r_n\right)$; *for Case (II),* (a') $\frac{1}{l}(X^{*'}P_Z X^* - \hat{\lambda}^*(\hat{\beta}^*)X^{*'}X^*) = \bar{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^\lambda n/l\right)$, (b') $\frac{1}{l}(X^{*'}P_Z\hat{\epsilon}^* - \hat{\lambda}^*(\hat{\beta}^*)X^{*'}\hat{\epsilon}^*) = O_{P^*}\left(1/\sqrt{l} + \delta_n^\beta + \delta_n^\lambda n/l\right)$, *where* $\bar{H}_{II,n} = (1-\lambda_n)\left(\hat{\Pi}'Z'Z\hat{\Pi}/l\right)$.

Lemma .3.7. *Suppose that Assumptions 1-2 hold. Suppose* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n \to 0$, *then in Case (I),* $\hat{\lambda}^*(\hat{\beta}^*) = \lambda^* + O_{P^*}\left(\frac{r_n}{n}(\delta_n^\beta)^2\right)$; *in Case (II),* $\hat{\lambda}^*(\hat{\beta}^*) = \lambda^* + O_{P^*}\left(\frac{l}{n}(\delta_n^\beta)^2\right)$.

Let

$$\hat{D}^*\left(\hat{\beta}^*\right) = \frac{\partial}{\partial\beta}\left(\frac{\left(y^* - X^*\hat{\beta}^*\right)' P_Z \left(y^* - X^*\hat{\beta}^*\right)}{2\left(y^* - X^*\hat{\beta}^*\right)'\left(y^* - X^*\hat{\beta}^*\right)}\right) = X^{*'} P_Z \varepsilon^*(\hat{\beta}^*) - \hat{\lambda}^*(\hat{\beta}^*) X^{*'}\varepsilon^*(\hat{\beta}^*),$$

where $\varepsilon^*(\hat{\beta}^*) = y^* - X^*\hat{\beta}^*$ and $\hat{\lambda}^*(\hat{\beta}^*) = \frac{\varepsilon^*(\hat{\beta}^*)' P_Z \varepsilon^*(\hat{\beta}^*)}{\varepsilon^*(\hat{\beta}^*)'\varepsilon^*(\hat{\beta}^*)}$.

Lemma .3.8. *Suppose that Assumptions 1-2 hold. Suppose* $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n^\beta \to 0$ *then for Case (I),*

$$-\frac{1}{r_n}\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = \bar{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n\right);$$

for Case (II),

$$-\frac{1}{l}\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = \bar{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n\right),$$

in probability, where $\bar{\beta}^*$ *lies between* $\hat{\beta}$ *and* $\hat{\beta}^*$.

Lemma .3.9. *Suppose that Assumptions 1-2 hold, then the following statements are true: in Case (I),* $\frac{1}{\sqrt{r_n}}\hat{D}^*\left(\hat{\beta}\right) = \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(1/\sqrt{r_n}\right)$*; in Case (II),* $\frac{1}{\sqrt{l}}\hat{D}^*\left(\hat{\beta}\right) = \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(1/\sqrt{l}\right)$*, in probability, where* $\tilde{V}^* = V^* - \epsilon^*\left(\frac{\sigma_{V\epsilon}^{b'}}{\sigma_{\epsilon\epsilon}^b}\right)$.

Proof of Theorem 3.3.1

Notice that the first-order conditions for the bootstrap analogue of LIML can be written as $\hat{D}^*(\hat{\beta}^*) = 0$ with

$$\begin{aligned}\hat{D}^*\left(\hat{\beta}^*\right) &= \frac{\partial}{\partial\beta}\left(\frac{\left(y^* - X^*\hat{\beta}^*\right)' P_Z \left(y^* - X^*\hat{\beta}^*\right)}{2\left(y^* - X^*\hat{\beta}^*\right)'\left(y^* - X^*\hat{\beta}^*\right)}\right) \\ &= X^{*'} P_Z \varepsilon^*(\hat{\beta}^*) - \hat{\lambda}^*(\hat{\beta}^*) X^{*'}\varepsilon^*(\hat{\beta}^*),\end{aligned}$$

where $\varepsilon^*(\hat{\beta}^*) = y^* - X^*\hat{\beta}^*$ and $\hat{\lambda}^*(\hat{\beta}^*) = \frac{\varepsilon^*(\hat{\beta}^*)' P_Z \varepsilon^*(\hat{\beta}^*)}{\varepsilon^*(\hat{\beta}^*)'\varepsilon^*(\hat{\beta}^*)}$. Expanding around $\hat{\beta}$ gives

$$0 = \hat{D}^*(\hat{\beta}) + \frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}(\hat{\beta}^* - \hat{\beta})$$

where $\bar{\beta}^*$ lies on the line joining $\hat{\beta}^*$ and $\hat{\beta}$;

Then, for Case (I) we have

$$\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) = -\left(\frac{1}{r_n}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1} \frac{1}{\sqrt{r_n}}\hat{D}^*(\hat{\beta})$$

and for Case (II),

$$\sqrt{l}(\hat{\beta}^* - \hat{\beta}) = -\left(\frac{1}{l}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1} \frac{1}{\sqrt{l}}\hat{D}^*(\hat{\beta}).$$

Lemma .3.8 establish the limit of $-\frac{1}{r_n}\left(\partial \hat{D}^*(\bar{\beta}^*)/\partial\beta\right)$ in Case (I) and the limit of $-\frac{1}{l}\left(\partial \hat{D}^*(\bar{\beta}^*)/\partial\beta\right)$ in Case (II). Also, by Lemma .3.9,

$$\frac{1}{\sqrt{r_n}}\hat{D}^*(\hat{\beta}) = \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$

in Case (I) , and

$$\frac{1}{\sqrt{l}}\hat{D}^*(\hat{\beta}) = \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

in Case (II).

To proceed, for Case (I) we let $W^*_{I,i} = \begin{pmatrix} \frac{1}{\sqrt{r_n}}(1-\lambda_n)\hat{\Pi}' Z_i \varepsilon^*_i \\ \frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}^*_i\varepsilon^*_i \end{pmatrix}$, and we check the conditions of Lemma A2 in Hansen, Hausman, and Newey (2008) hold with $W_i = W^*_{I,i}$, $v_i = \tilde{V}^*_i$ and $u_i = \epsilon^*_i$, where $\tilde{V}^*_i \equiv V^*_i - \varepsilon^*_i\left(\frac{\sigma^{b\prime}_{V\epsilon}}{\sigma^b_{\varepsilon\varepsilon}}\right)$. We need to show that

$$\begin{cases} \sum_{i=1}^{n} E^{*}\left(\left\|\frac{1}{\sqrt{r_n}}(1-\lambda_n)\hat{\Pi}' Z_i \varepsilon_i^{*}\right\|^{4}\right) \to_P 0 \\ \sum_{i=1}^{n} E^{*}\left(\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^{*} \varepsilon_i^{*}\right\|^{4}\right) \to_P 0 \end{cases}$$

For the first term, notice that

$$\begin{aligned} & E^{*}\sum_{i=1}^{n}\left\|\frac{1}{\sqrt{r_n}}\hat{\Pi}' Z_i \epsilon_i^{*}\right\|^{4} \\ = \; & E^{*}\sum_{i=1}^{n}\left\|\frac{1}{\sqrt{r_n}}\Pi' Z_i \epsilon_i^{*} + \frac{1}{\sqrt{r_n}} V' Z (Z'Z)^{-1} Z_i \epsilon_i^{*}\right\|^{4} \\ \leq \; & C_1 \sum_{i=1}^{n}\left\|\frac{1}{\sqrt{r_n}}\Pi' Z_i\right\|^{4} E^{*}\left(\epsilon_i^{*4}\right) + C_1 \sum_{i=1}^{n}\left\|\frac{1}{\sqrt{r_n}} V' Z (Z'Z)^{-1} Z_i\right\|^{4} E^{*}\left(\epsilon_i^{*4}\right) \\ \equiv \; & D_1 + D_2 \end{aligned}$$

where the inequality follows from Minkowski inequality. Note that $E^{*}\left(\epsilon_i^{*4}\right)$ is bounded in probability from similar arguments as in Lemma .3.1, therefore we have

$$D_1 = O_P(1) C_1 \left(\frac{1}{r_n^2}\sum_{i=1}^{n} \| \Pi' Z_i \|^{4}\right) \to^{P} 0$$

by Assumption 2. Similarly, we have for D_2

$$D_2 \;=\; O_P(1) C_1 \left(\frac{1}{r_n^2}\sum_{i=1}^{n} \left\| V' Z (Z'Z)^{-1} Z_i \right\|^{4}\right)$$

Let $w = (w_1, ..., w_n)'$ be an arbitrary column of V, then by Marcinkiewicz-Zygmund inequality,

$$E\left[\left|w' Z (Z'Z)^{-1} Z_i\right|^{4}\right] = E\left[\left|\sum_{j=1}^{n} w_j P_{ji}\right|^{4}\right] \leq CE\left[\left|\sum_{j=1}^{n} w_j^2 P_{ji}^2\right|^{2}\right]$$

By $P_{ii} \leq 1$ it follows that $(P_{ii})^2 \leq P_{ii}$. Also, $f(r) = r^2$ is a convex function of r. Then by Jensen's inequality and $\sum_{j=1}^n (P_{ji})^2 = P_{ii}$ we have

$$E\left[\left|\sum_{j=1}^n w_j^2 P_{ji}^2\right|^2\right] \leq (P_{ii})^2 E\left[\left|\sum_{j=1}^n w_j^2 (P_{ji})^2 / P_{ii}\right|^2\right] = P_{ii}\sum_{j=1}^n E\left(|w_j|^4\right)\left(\frac{(P_{ji})^2}{P_{ii}}\right) \leq CP_{ii}$$

Combining the last two equations gives $E\left[\|w'Z(Z'Z)^{-1}Z_i\|^4\right] \leq CP_{ii}$. Therefore

$$\frac{1}{r_n^2}\sum_{i=1}^n E\left(\left\|w'Z(Z'Z)^{-1}Z_i\right\|^4\right) \leq C\left(\frac{\sum_{i=1}^n P_{ii}}{r_n^2}\right) = C\left(\frac{l}{r_n^2}\right) \to 0$$

by $\sum_{i=1}^n P_{ii} = l$ and $\sqrt{l}/r_n \to 0$, then we obtain $\frac{1}{r_n^2}\sum_{i=1}^n \|w'Z(Z'Z)^{-1}Z_i\|^4 \to^P 0$ by Markov inequality. The conclusion for V follows by showing the result for each column. Then, we obtain $D_2 \to^P 0$, and the result for the first term follows.

Also, for the second term $\sum_{i=1}^n E^*\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right\|^4$, we have by $E^*\left(\epsilon_i^{*8}\right)$ and $E^*\left(V_i^{*8}\right)$ being bounded in probability that

$$\sum_{i=1}^n E^*\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\epsilon_i^*\right\|^4 = O_P(1)\left(\frac{\sum_{i=1}^n P_{ii}^4 + n\lambda_n^4}{l^2}\right) \leq O_P(1)\left(\frac{1}{l}+\frac{\lambda_n^2}{n}\right) \to^P 0$$

as required. Then for case (I),

$$\begin{aligned}
&\sum_{i=1}^n E^*\left(W_{I,i}^* W_{I,i}^{*\prime}\right) \\
=& \begin{pmatrix} \sum_{i=1}^n \frac{1}{r_n}(1-\lambda_n)^2 \hat{\Pi}' Z_i Z_i' \hat{\Pi} E^*\left(\varepsilon_i^{*2}\right) & \sum_{i=1}^n \frac{1-\lambda_n}{\sqrt{l r_n}}(P_{ii}-\lambda_n)\hat{\Pi}' Z_i E^*\left(\varepsilon_i^{*2}\tilde{V}_i^{*\prime}\right) \\ E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\right)\sum_{i=1}^n \frac{1-\lambda_n}{\sqrt{l r_n}}(P_{ii}-\lambda_n) Z_i'\hat{\Pi} & \sum_{i=1}^n \frac{1}{l}(P_{ii}-\lambda_n)^2 E^*\left(\varepsilon_i^{*2}\tilde{V}_i^*\tilde{V}_i^{*\prime}\right) \end{pmatrix} \\
\to^P & \begin{pmatrix} (1-\lambda)\sigma_{\epsilon\epsilon}\bar{H}_I & (1-\lambda)\bar{A}' \\ (1-\lambda)\bar{A} & (\phi-\lambda)\left(\sigma_{\epsilon\epsilon}\bar{\Sigma}_{\tilde{V}\tilde{V}} + \bar{B}\right) \end{pmatrix} \equiv \bar{\Psi}_I,
\end{aligned}$$

where

$$\begin{aligned}
\bar{H}_I &= (1-\lambda)\left(Q+\gamma\Sigma_{VV}\right); \\
\bar{\Sigma}_{\tilde{V}\tilde{V}} &= \Sigma_{\tilde{V}\tilde{V}} - \lambda\left(\Sigma_{VV} + (\lambda-2)\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right); \\
\bar{A} &= (1-\lambda)A, q = \frac{\sigma_{V\epsilon}}{\sigma_{\epsilon\epsilon}}; \\
\bar{B} &= (1-2\lambda+\lambda\phi)(\phi-\lambda)B + 2\lambda(\phi-\lambda)^2 E\left(\epsilon_i^3\tilde{V}_i\right)q' + \lambda(\phi-\lambda)^2 q\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right)q'.
\end{aligned}$$

These results follow from

$$\begin{aligned}
&\sum_{i=1}^{n}\frac{1}{r_n}(1-\lambda_n)^2\hat{\Pi}'Z_iZ_i'\hat{\Pi}E^*\left(\varepsilon_i^{*2}\right) \\
=\ & (1-\lambda_n)^2\left(\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n}\right)\left(\frac{\Pi'Z'Z\Pi}{r_n} + \frac{\Pi'Z'V}{r_n} + \frac{V'Z\Pi}{r_n} + \frac{V'P_ZV}{r_n}\right) \\
=\ & (1-\lambda_n)^2\left(\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n}\right)\left\{\frac{\Pi'Z'Z\Pi}{r_n} + O_P\left(\frac{1}{\sqrt{r_n}}\right) + O_P\left(\frac{1}{\sqrt{r_n}}\right) + \left(\frac{l}{r_n}\right)\left[\Sigma_{VV} + O_P\left(\frac{1}{\sqrt{l}}\right)\right]\right\} \\
\to^P\ & (1-\lambda)^2\sigma_{\epsilon\epsilon}\left(Q+\gamma\Sigma_{VV}\right) \equiv (1-\lambda)\sigma_{\epsilon\epsilon}\bar{H}_I
\end{aligned}$$

since $\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n} \to^P \sigma_{\epsilon\epsilon}$, $\frac{\Pi'Z'Z\Pi}{r_n} \to Q$, and $\frac{\Pi'Z'V}{r_n} = O_P\left(\frac{1}{\sqrt{r_n}}\right)$.

For the off-diagonal term $\sum_{i=1}^{n}\frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\hat{\Pi}'Z_iE^*\left(\varepsilon_i^{*2}\tilde{V}_i^{*\prime}\right)$, we note that

$$\begin{aligned}
E^*\left(\epsilon_i^{*2}\tilde{V}_i^{*\prime}\right) &= E^*\left(\epsilon_i^{*2}V_i^{*\prime}\right) - E^*\left(\epsilon_i^{*3}\right)\left(\frac{\sigma_{V\epsilon}^{b\prime}}{\sigma_{\epsilon\epsilon}^{b}}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i^2\hat{V}_i' - \frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i^3\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i\hat{V}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\epsilon}_i^2\right)^{-1}
\end{aligned}$$

Let $\hat{a} = (\hat{\epsilon}_1^2 - \sigma_{\epsilon\epsilon}, ..., \hat{\epsilon}_n^2 - \sigma_{\epsilon\epsilon})'$ and $a = (\epsilon_1^2 - \sigma_{\epsilon\epsilon}, ..., \epsilon_n^2 - \sigma_{\epsilon\epsilon})'$. Then, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} \tilde{\epsilon}_i^2 \hat{V}_i' - \frac{a'\hat{V}}{n} &= \frac{1}{n}\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})^2 \hat{V}_i' - \frac{a'\hat{V}}{n} \\ &= \frac{(\hat{a}-a)'\hat{V}}{n} - 2\bar{\hat{\epsilon}}\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i \hat{V}_i'\right) + \bar{\hat{\epsilon}}^2 \left(\frac{1}{n}\sum_{i=1}^{n} \hat{V}_i\right)' \\ &= o_P(1) + o_P(1) + o_P(1) = o_P(1) \end{aligned}$$

where $\bar{\hat{\epsilon}} = n^{-1}\sum_{i=1}^{n} \hat{\epsilon}_i$. The results follows by noting that $\frac{\|\hat{a}-a\|^2}{n} = O_P\left(\frac{1}{r_n}\right)$, $\frac{\hat{V}'\hat{V}}{n} = \frac{V'M_Z V}{n} = (1-\lambda)\Sigma_{VV} + o_P(1) = O_P(1)$, $\bar{\hat{\epsilon}} \to^P E(\epsilon_i) = 0$, $n^{-1}\sum_{i=1}^{n} \hat{\epsilon}_i \hat{V}_i' = O_P(1)$ and $n^{-1}\sum_{i=1}^{n} \hat{V}_i = O_P(1)$. Also, by using arguments similar to those in Lemma A9 of Hansen, Hausman, and Newey (2008), we obtain $\frac{a'\hat{V}}{n} \to^P (1-\lambda)E\left(\epsilon_i^2 V_i'\right)$. Thus, it follows that $E^*\left(\epsilon_i^{*2} V_i^{*\prime}\right) \to^P (1-\lambda)E\left(\epsilon_i^2 V_i'\right)$. For the second term in $E^*\left(\epsilon_i^{*2} \tilde{V}_i^{*\prime}\right)$, we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \tilde{\epsilon}_i^3 \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\epsilon}_i \hat{V}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{\epsilon}_i^2\right)^{-1} = \frac{1}{n}\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})^3 \left(\frac{1}{n}\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})\hat{V}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} (\hat{\epsilon}_i - \bar{\epsilon})^2\right)^{-1} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^3 + o_P(1)\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i \hat{V}_i' + o_P(1)\right)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 + o_P(1)\right)^{-1} \to^P (1-\lambda)E\left(\epsilon_i^3\right)\left(\frac{\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right) \end{aligned}$$

where the second equality follows by $\bar{\epsilon} \to^P 0$, and the convergence in probability follows by $\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^3 \to^P E\left(\epsilon_i^3\right)$, $\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i \hat{V}_i' \to^P (1-\lambda)E\left(\epsilon_i V_i'\right)$, and $\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 \to^P E\left(\epsilon_i^2\right)$. Therefore, for the standard residual bootstrap, we obtain $E^*\left(\epsilon_i^{*2} \tilde{V}_i^{*\prime}\right) \to^P (1-\lambda)E\left(\epsilon_i^2 V_i'\right) - (1-\lambda)E\left(\epsilon_i^3\right)\left(\frac{\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right) = (1-\lambda)E\left(\epsilon_i^2 \tilde{V}_i'\right)$.

For $E^*\left(\epsilon_i^{*2} \tilde{V}_i^* \tilde{V}_i^{*\prime}\right)$, notice that $E^*\left(\epsilon_i^{*2} \tilde{V}_i^* \tilde{V}_i^{*\prime}\right) = \sigma_{\epsilon\epsilon}^b E^*\left(\tilde{V}_i^* \tilde{V}_i^{*\prime}\right) + E^*\left(\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right) \tilde{V}_i^* \tilde{V}_i^{*\prime}\right)$.

For the first term, by the bootstrap DGP, we have

$$\begin{aligned} E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right) &= E^*\left\{\left(V_i^* - \epsilon_i^*\left(\frac{\sigma_{V\epsilon}^{b'}}{\sigma_{\epsilon\epsilon}^b}\right)\right)\left(V_i^* - \epsilon_i^*\left(\frac{\sigma_{V\epsilon}^{b'}}{\sigma_{\epsilon\epsilon}^b}\right)'\right)\right\} \\ &= \frac{\hat{V}'\hat{V}}{n} - \left(\frac{\hat{V}'\tilde{\epsilon}}{n}\right)\left(\frac{\tilde{\epsilon}'\hat{V}}{n}\right)\left(\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n}\right)^{-1} \to^P (1-\lambda)\Sigma_{VV} - (1-\lambda)^2\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}} \end{aligned}$$

which follows from $\hat{V}'\hat{V}/n \to^P (1-\lambda)\Sigma_{VV}$, $\hat{V}'\tilde{\epsilon}/n \to^P (1-\lambda)\sigma_{V\epsilon}$, and $\tilde{\epsilon}'\tilde{\epsilon}/n \to^P \sigma_{\epsilon\epsilon}$. Also, $\sigma_{\epsilon\epsilon}^b = \frac{\tilde{\epsilon}'\tilde{\epsilon}}{n} \to^P \sigma_{\epsilon\epsilon}$. For the second term $E^*\left(\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right)\tilde{V}_i^*\tilde{V}_i^{*'}\right)$, we let $\tilde{\omega}$ be a column of $\tilde{V}$ and ω be a column of V, also let $\tilde{\omega}^*$ be a column of $\tilde{V}^*$ and ω^* be a column of V^*. By the standard bootstrap scheme,

$$E^*\left(\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right)\tilde{\omega}_i^{*2}\right) = \frac{1}{n}\sum_{i=1}^n\left\{(\tilde{\epsilon}_i^2 - \sigma_{\epsilon\epsilon}^b)\left[\hat{\omega}_i - \tilde{\epsilon}_i\left(\frac{\sigma_{\epsilon\omega}^b}{\sigma_{\epsilon\epsilon}^b}\right)\right]^2\right\}$$

where $\hat{\omega} = M_Z\omega$ and $\sigma_{\epsilon\omega}^b = \frac{\tilde{\epsilon}'\hat{\omega}}{n}$. Also note that

$$\hat{\omega} - \tilde{\epsilon}\left(\frac{\sigma_{\epsilon\omega}^b}{\sigma_{\epsilon\epsilon}^b}\right) = M_Z\omega - \tilde{\epsilon}\left(\frac{\tilde{\epsilon}'M_Z\omega}{n}\right)\left(\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n}\right)^{-1} = M_Z\tilde{\omega} + M_Z\epsilon q_\omega - (1-\lambda_n)\epsilon q_\omega + o_P(1)$$

where $\tilde{\omega} = \omega - \epsilon q_\omega$ and $q_\omega \equiv \frac{\sigma_{\epsilon\omega}}{\sigma_{\epsilon\epsilon}}$; the second equality follows by the fact that $\tilde{\epsilon} = \epsilon + o_P(1)$, $\frac{\tilde{\epsilon}'M_Z\omega}{n} = (1-\lambda_n)\sigma_{\epsilon w} + o_P(1)$ and $\frac{\tilde{\epsilon}'\tilde{\epsilon}}{n} = \sigma_{\epsilon\epsilon} + o_P(1)$. Denote $A = diag(a_1, ..., a_n)$, then

$$\begin{aligned} & E^*\left(\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right)\tilde{\omega}_i^{*2}\right) \\ = & \frac{1}{n}\left(M_Z\tilde{\omega} + M_Z\epsilon q_\omega - (1-\lambda_n)\epsilon q_\omega + o_P(1)\right)' A\left(M_Z\tilde{\omega} + M_Z\epsilon q_\omega - (1-\lambda_n)\epsilon q_\omega + o_P(1)\right) \\ = & \frac{\tilde{\omega}'M_ZAM_Z\tilde{\omega}}{n} + 2q_\omega\left(\frac{\epsilon'M_ZAM_Z\tilde{\omega}}{n}\right) - 2(1-\lambda_n)q_\omega\left(\frac{\epsilon'AM_Z\tilde{\omega}}{n}\right) + q_\omega^2\left(\frac{\epsilon'M_ZAM_Z\epsilon}{n}\right) \\ & -2(1-\lambda_n)q_\omega^2\left(\frac{\epsilon'AM_Z\epsilon}{n}\right) + (1-\lambda)^2q_\omega^2\left(\frac{\epsilon'A\epsilon}{n}\right) + o_P(1) \\ \to^P & (1-2\lambda+\lambda\phi)E\left(\left(\epsilon_i^2 - \sigma_{\epsilon\epsilon}\right)\tilde{\omega}_i^2\right) + 2q_\omega\lambda(\phi-\lambda)E\left(\epsilon_i^3\tilde{\omega}_i\right) + q_\omega^2\lambda(\phi-\lambda)\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right) \end{aligned}$$

by showing that

$$\frac{\tilde{\omega}' M_Z A M_Z \tilde{\omega}}{n} \to^P (1 - 2\lambda + \lambda\phi) E\left(\left(\epsilon_i^2 - \sigma_{\epsilon\epsilon}\right)\tilde{\omega}_i^2\right);$$

$$\frac{\epsilon' M_Z A M_Z \tilde{\omega}}{n} \to^P (1 - 2\lambda + \lambda\phi) E\left(\epsilon_i^3 \tilde{\omega}_i\right);$$

$$\frac{\epsilon' M_Z A M_Z \epsilon}{n} \to^P (1 - 2\lambda + \lambda\phi)\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right);$$

$\frac{\epsilon' A M_Z \tilde{\omega}}{n} \to^P (1-\lambda) E\left(\epsilon_i^3 \tilde{w}_i\right)$; $\frac{\epsilon' A M_Z \epsilon}{n} \to^P (1-\lambda)\left(E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2\right)$; and $\frac{\epsilon' A \epsilon}{n} \to^P E\left(\epsilon_i^4\right) - (\sigma_{\epsilon\epsilon})^2$ using similar arguments as in Lemma A10 and A11 in Hansen, Hausman, and Newey (2008). We apply this result to each component of $\tilde{V}$ and the result for $E^*\left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b)\tilde{V}_i^* \tilde{V}_i^{*\prime}\right)$ follows.

Define $U_I^* = \left(\sum_{i=1}^n W_{I,i}^*, \sum_{i\neq j}^n \frac{1}{\sqrt{l}} \tilde{V}_i^* P_{ij} \varepsilon_j^*\right)'$, then we have by Lemma A2 of Hansen, Hausman, and Newey (2008) that

$$U_I^* \to^{d^*} N\left(0, \begin{pmatrix} \Psi_I & 0 \\ 0 & (1-\phi)\sigma_{\epsilon\epsilon}\bar{\Sigma}_{\tilde{V}\tilde{V}} \end{pmatrix}\right), \text{ in probability}$$

Also define $F_{I,n} = \left(1 \ \sqrt{\frac{l}{r_n}} \ \sqrt{\frac{l}{r_n}}\right)$, then $F_{I,n} \to F_0 \equiv \left(1 \ \sqrt{\gamma} \ \sqrt{\gamma}\right)$ and by Lemma .3.9

$$\begin{aligned} \frac{1}{\sqrt{r_n}} \hat{D}^*(\hat{\beta}) &= \frac{1}{\sqrt{r_n}} \left((1-\lambda_n) Z\hat{\Pi} + P_Z \tilde{V}^* - \lambda_n \tilde{V}^*\right)' \epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right), \text{ in probability} \\ &= F_I \times U^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right), \text{ in probability} \\ &\to^{d^*} N(0, \bar{\Upsilon}_I), \text{ in probability} \end{aligned}$$

with $\bar{\Upsilon}_I = F_0 \begin{pmatrix} \Psi_I & 0 \\ 0 & (1-\phi)\sigma_{\epsilon\epsilon}\bar{\Sigma}_{\tilde{V}\tilde{V}} \end{pmatrix} F_0' = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\bar{H}_I + \gamma\bar{\Sigma}_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\bar{A} + \bar{A}'\right\} + \gamma\bar{B}$; together with the result in Lemma .3.8, we obtain that in case (I),

$$\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) = \left(\frac{1}{r_n} \frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta} \right)^{-1} \frac{1}{\sqrt{r_n}} \hat{D}^*(\hat{\beta}) \to^{d^*} N\left(0, \bar{\Lambda}_I\right), \text{ in probability.}$$

where $\bar{\Lambda}_I = \bar{H}_I^{-1} \bar{\Upsilon}_I \bar{H}_I^{-1}$.

Now we turn to case (II). Let

$$W^*_{II,i} = \left(\frac{1}{\sqrt{l}} (1-\lambda_n) \hat{\Pi}' Z_i \varepsilon^*_i, \frac{1}{\sqrt{l}} (P_{ii} - \lambda_n) \tilde{V}^*_i \varepsilon^*_i \right)',$$

Using arguments similar to Case (I), we obtain

$$\sum_{i=1}^n E^* \left(\left\| \frac{1}{\sqrt{l}} (1-\lambda_n) \hat{\Pi}' Z_i \epsilon^*_i \right\|^4 \right) \to^P 0,$$

and

$$\begin{aligned} & \sum_{i=1}^n E^* \left(W^*_{II,i} W^{*\prime}_{II,i} \right) \\ = & \begin{pmatrix} \sum_{i=1}^n \frac{(1-\lambda_n)^2}{l} \hat{\Pi}' Z_i Z_i' \hat{\Pi} E^* \left(\varepsilon_i^{*2} \right) & \sum_{i=1}^n \frac{1-\lambda_n}{l} (P_{ii} - \lambda_n) \hat{\Pi}' Z_i E^* \left(\varepsilon_i^{*2} \tilde{V}_i^{*\prime} \right) \\ E^* \left(\varepsilon_i^{*2} \tilde{V}_i^* \right) \sum_{i=1}^n \frac{1-\lambda_n}{l} (P_{ii} - \lambda_n) Z_i' \hat{\Pi} & \sum_{i=1}^n \frac{1}{l} (P_{ii} - \lambda_n)^2 E^* \left(\varepsilon_i^{*2} \tilde{V}_i^* \tilde{V}_i^{*\prime} \right) \end{pmatrix} \\ \to_P & \begin{pmatrix} (1-\lambda) \sigma_{\epsilon\epsilon} \bar{H}_{II} & 0 \\ 0 & (\phi - \lambda) \left(\sigma_{\epsilon\epsilon} \bar{\Sigma}_{\tilde{V}\tilde{V}} + \bar{B} \right) \end{pmatrix} \equiv \bar{\Psi}_{II}, \end{aligned}$$

where $\bar{H}_{II} = (1-\lambda)\Sigma_{VV}$. Notice that the off-diagonal terms converges in probability to zero because in Case (II), $\sum_{i=1}^n \left(\frac{(P_{ii} - \lambda_n) \hat{\Pi}' Z_i}{l} \right) \to^P 0$.

Define $U^*_{II} = \left(\sum_{i=1}^n W^*_{II,i}, \sum_{i \neq j}^n \frac{1}{\sqrt{l}} \tilde{V}^*_i P_{ij} \varepsilon^*_i \right)'$, then we have by Lemma A2 of Hansen, Hausman, and Newey (2008) that

$$U^*_{II} \to^{d^*} N \left(0, \begin{pmatrix} \bar{\Psi}_{II} & 0 \\ 0 & (1-\phi) \sigma_{\epsilon\epsilon} \bar{\Sigma}_{\tilde{V}\tilde{V}} \end{pmatrix} \right) \text{ in probability.}$$

Also define $F_{II} = [1\ 1\ 1]$, then

$$\begin{aligned}
\frac{1}{\sqrt{l}}\hat{D}^*(\hat{\beta}) &= \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \\
&= F_{II} \times U_{II}^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \\
&\to^{d^*} N\left(0, \bar{\Upsilon}_{II}\right), \text{ in probability}
\end{aligned}$$

where $\bar{\Upsilon}_{II} = F_{II}\begin{pmatrix}\Psi_{II} & 0 \\ 0 & (1-\phi)\sigma_{\epsilon\epsilon}\bar{\Sigma}_{\tilde{V}\tilde{V}}\end{pmatrix}F_{II}' = \sigma_{\epsilon\epsilon}\left((1-\lambda)\bar{H}_{II} + (1-\phi)\bar{\Sigma}_{\tilde{V}\tilde{V}}\right)$. Together with Lemma .3.7, this leads to the result that in case (II),

$$\sqrt{l}\left(\hat{\beta}^* - \hat{\beta}\right) = \left(\frac{1}{l}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{l}}\hat{D}^*(\hat{\beta}) \to^{d^*} N\left(0, \bar{\Lambda}_{II}\right) \text{ in probability}$$

where $\bar{\Lambda}_{II} = \bar{H}_{II}^{-1}\bar{\Upsilon}_{II}\bar{H}_{II}^{-1}$. ■

Proof of Corollary 3.3.1

By Theorem 3.3.1, we have when $l/r_n \to 0$

$$\sqrt{r_n}(\hat{\beta}^* - \hat{\beta}) \to^{d^*} N(0, \sigma_{\epsilon\epsilon}Q^{-1})$$

in probability, which is the same as the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta_0)$ under $l/r_n \to 0$. The result therefore follows by Polya's Theorem. ■

Now we give the proofs for the Lemmas.

Proof of Lemma .3.1

(a) Let $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^n \epsilon_i$, $\bar{V} = \frac{1}{n}\sum_{i=1}^n V_i$, $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

Using Minkowski and Cauchy-Schwartz inequalities, we obtain

$$\begin{aligned} E^{*}\left(\epsilon_i^{*8}\right) &= \frac{1}{n}\sum_{i=1}^{n}\left(\epsilon_i - \bar{\epsilon} - (X_i - \bar{X})'(\hat{\beta} - \beta)\right)^8 \\ &\leq C_1\left\{\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i - \bar{\epsilon}|^8 + \frac{1}{n}\sum_{i=1}^{n}|(X_i - \bar{X})'(\hat{\beta} - \beta)|^8\right\} \\ &\leq C_2\left\{\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i - \bar{\epsilon}|^8 + \left\|\hat{\beta} - \beta\right\|^8 \frac{1}{n}\sum_{i=1}^{n}\left\|X_i - \bar{X}\right\|^8\right\} \end{aligned}$$

for large enough constants C_1 and C_2. Using Minkowski inequality again, we obtain

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left\|X_i - \bar{X}\right\|^8 &= n^{-1}\sum_{i=1}^{n}\left\|\Pi'\left(Z_i - \bar{Z}\right) + \left(V_i - \bar{V}\right)\right\|^8 \\ &\leq C\left\{n^{-1}\sum_{i=1}^{n}\left\|\Pi'\left(Z_i - \bar{Z}\right)\right\|^8 + n^{-1}\sum_{i=1}^{n}\left\|V_i - \bar{V}\right\|^8\right\} = O_P(1) \end{aligned}$$

which follows from Z being fixed and $n^{-1}\sum_{i=1}^{n}\left\|V_i - \bar{V}\right\|^8 = O_P(1)$ since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|V_i - \bar{V}\right\|^8 \leq C\left\{\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 + \left\|\bar{V}\right\|^8\right\} \to^P C\left\{E\|V_i\|^8 + \|E\left(V_i\right)\|^8\right\}$$

and $\|E\left(V_i\right)\| \leq E\|V_i\| \leq \left(E\|V_i\|^8\right)^{1/8}$ by Jensen's inequality. Similarly, we have $n^{-1}\sum_{i=1}^{n}\|\epsilon_i - \bar{\epsilon}\|^8 = O_P(1)$. Also, $\hat{\beta} - \beta \to^P 0$ under Assumption 1 and 2, $E^{*}\left(\epsilon_i^{*8}\right)$ is thus bounded in probability.

(b) Similarly,

$$\begin{aligned} E^*\left(\|V_i^*\|^8\right) &= \frac{1}{n}\sum_{i=1}^{n}\left\|V_i - \left(\hat{\Pi}-\Pi\right)' Z_i\right\|^8 \\ &\leq C_1\left\{\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 + \frac{1}{n}\sum_{i=1}^{n}\left\|\left(\hat{\Pi}-\Pi\right)' Z_i\right\|^8\right\} \\ &= C_1\left\{\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 + \frac{1}{n}\sum_{i=1}^{n}\left\|V'Z\left(Z'Z\right)^{-1}Z_i\right\|^8\right\} \end{aligned}$$

Since by Assumption 1, $\frac{1}{n}\sum_{i=1}^{n}\|V_i\|^8 = O_P(1)$, it suffices to show that $\frac{1}{n}\sum_{i=1}^{n}\left\|V'Z\left(Z'Z\right)^{-1}Z_i\right\|^8 = O_P(1)$. We are going to show that this holds for each element of the vector $V'Z\left(Z'Z\right)^{-1}Z_i$. Let $w = (w_1,\ldots,w_n)$ an arbitrary column of V. Then, by Marcinkiewicz-Zygmund inequality

$$E\left|w'Z\left(Z'Z\right)^{-1}Z_i\right|^8 = E\left|\sum_{j=1}^{n} w_j P_{ji}\right|^8 \leq C_2 E\left|\sum_{j=1}^{n} w_j^2 P_{ji}^2\right|^4$$

Also notice that $\left(\sum_{j=1}^{n} P_{ji}^2\right)^{-4} \geq 1$, then by Jensen's inequality,

$$\begin{aligned} E\left|\sum_{j=1}^{n} w_j^2 P_{ji}^2\right|^4 &\leq E\left|\sum_{j=1}^{n} w_j^2 P_{ji}^2 / \left(\sum_{j=1}^{n} P_{ji}^2\right)\right|^4 \leq E\left[\sum_{j=1}^{n} w_j^8 P_{ji}^2 / \left(\sum_{j=1}^{n} P_{ji}^2\right)\right] \\ &\leq \sum_{j=1}^{n} E\left(w_j^8\right) P_{ji}^2 / \left(\sum_{j=1}^{n} P_{ji}^2\right) \leq C_3 \end{aligned}$$

Thus, $E\left\|w'Z(Z'Z)^{-1}Z_i\right\|^8 = O(1)$. Applying this result to each column of V, we obtain $E\left\|V'Z(Z'Z)^{-1}Z_i\right\|^8 = O(1)$. Thus, by Markov inequality, $\frac{1}{n}\sum_{i=1}^{n}\left\|V'Z(Z'Z)^{-1}Z_i\right\|^8 = O_P(1)$, and the conclusion of part (b) follows. ■

Proof of Lemma .3.2

To prove part(a), note that it suffices to prove that $V^{*(g)\prime} P_Z \epsilon^* / l = \sigma_{V\epsilon}^{b(g)} +$

$O_{P^*}\left(1/\sqrt{l}\right)$ as $n \to \infty$, where $V^{*(g)}$ denoted the g-th column of V^*, so that $V^{*(g)'}P_Z\epsilon^*/l$ is the g-th element of $V^{*'}P_Z\epsilon^*/l$, and where $\sigma_{V\epsilon}^{b(g)}$ denotes the g-th element of $\sigma_{V\epsilon}^b$, $g = 1, ..., k$.

From the bootstrap DGP, we have

$$E^*\left(\frac{V^{*(g)'}P_Z\epsilon^*}{l}\right) = \left(\frac{1}{l}\right) trace\left(P_Z E^*\left(\epsilon^* V^{*(g)'}\right)\right) = \left(\frac{\sigma_{V\epsilon}^{b(g)}}{l}\right) trace(P_Z) = \sigma_{V\epsilon}^{b(g)}$$

because $E^*\left(\epsilon_i^* V_j^{*(g)}\right) = E^*\left(\epsilon_i^*\right) E^*\left(V_j^{*(g)}\right) = 0$ for $i \neq j$ by the property of i.i.d. bootstrap.

Furthermore, note that

$$\begin{aligned} & E^*\left(\frac{V^{*(g)'}P_Z\epsilon^*}{l} - \sigma_{V\epsilon}^{b(g)}\right)^2 \\ = \; & \frac{1}{l^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} P_{ij}P_{kl}E^*\left(V_{ig}^*\epsilon_j^* V_{kg}^*\epsilon_l^*\right) - \left(\frac{2\sigma_{V\epsilon}^{(g)b}}{l}\right)\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}E^*\left(V_{ig}^*\epsilon_j^*\right) + \left(\sigma_{V\epsilon}^{b(g)}\right)^2 \\ = \; & \frac{1}{l^2}E^*\left(V_{ig}^{*2}\epsilon_i^{*2}\right)\left[\sum_{i=1}^{n}(P_{ii})^2\right] + \frac{2}{l^2}\left(\Sigma_{VV}^{b(g,g)}\sigma_{\epsilon\epsilon}^b\right)\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}(P_{ij})^2\right] \\ & + \left\{\frac{2}{l^2}\left(\sigma_{V\epsilon}^{b(g)}\right)^2\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1}\left(P_{ii}P_{jj} + (P_{ij})^2\right)\right] - \left(\sigma_{V\epsilon}^{b(g)}\right)^2\right\} \\ \equiv \; & L_1 + L_2 + L_3 \end{aligned}$$

The second equality follows from noting that $E^*\left(V_{ig}^*\epsilon_j^* V_{kg}^*\epsilon_l^*\right)$ equals zero except in the case where either $(i = j = k = l)$ or $(i = k, j = l)$ or $(i = j, k = l)$ or $(i = l, j = k)$ and from using $\sum_{i=1}^{n} P_{ii} = l$.

Focusing on L_1 first, notice that

$$\begin{aligned} L_1 &\leq \frac{1}{l^2}\left(\frac{1}{n}E^*\left(V_{ig}^{*4}\right)\right)^{1/2}\left(E^*\left(\epsilon_i^{*4}\right)\right)^{1/2}\left[\sum_{i=1}^{n}\left(P_{ii}\right)^2\right] \\ &\leq \frac{1}{l}\left(\frac{1}{n}E^*\left(V_{ig}^{*4}\right)\right)^{1/2}\left(E^*\left(\epsilon_i^{*4}\right)\right)^{1/2} = O_P\left(\frac{1}{l}\right) \end{aligned}$$

where the first inequality follows from the Cauchy-Schwartz inequality, and the second inequality follows from using $\sum_{i=1}^{n}\left(P_{ii}\right)^2 \leq \sum_{i=1}^{n} P_{ii} = l$. The last equality follows from using the same arguments as in Lemma .3.1.

Next, for L_2, we have

$$L_2 \leq \left(\frac{\Sigma_{VV}^{b(g,g)}\sigma_{\epsilon\epsilon}^{b}}{l^2}\right)\left[\sum_{i=1}^{n}\left(P_{ii}\right)^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(P_{ij})^2\right] = \frac{\Sigma_{VV}^{b(g,g)}\sigma_{\epsilon\epsilon}^{b}}{l} = O_P\left(\frac{1}{l}\right)$$

because $\sum_{i=1}^{n}\left(P_{ii}\right)^2 + 2\sum_{i=2}^{n}\sum_{j=1}^{i-1}(P_{ij})^2 = Tr(P_Z'P_Z) = Tr(P_Z) = l$ given that P_Z is symmetric and idempotent.

Finally, for L_3, we note that

$$\begin{aligned} |L_3| &= \left|\frac{(\sigma_{V\epsilon}^{b(g)})^2}{l^2}\left[(Tr(P_Z))^2 + Tr(P_Z'P_Z) - 2\sum_{i=1}^{n}(P_{ii})^2\right] - \left(\sigma_{V\epsilon}^{b(g)}\right)^2\right| \\ &= \left|\frac{\left(\sigma_{V\epsilon}^{b(g)}\right)^2}{l^2}\left(l - 2\sum_{i=1}^{n}\left(P_{ii}\right)^2\right)\right| \leq \frac{\left(\sigma_{V\epsilon}^{b(g)}\right)^2}{l} + \frac{2\left(\sigma_{V\epsilon}^{b(g)}\right)^2\sum_{i=1}^{n}P_{ii}}{l^2} = O_P\left(\frac{1}{l}\right) \end{aligned}$$

Therefore, we obtain $E^*\left(\frac{V^{*(g)\prime}P_Z\epsilon^*}{l} - \sigma_{V\epsilon}^{b(g)}\right)^2 = O_P\left(\frac{1}{\sqrt{l}}\right)$.

But, for any T^* such that $Var^*\left(T^*\right) = O_P(1/l)$, where Var^* denotes the variance computed under P^*, by the Tchebychev's inequality, we have for any $\delta > 0$ and any fixed $M_\delta > 0$,

$$P^*\left(\mid \sqrt{l}T^* \mid > M_\delta\right) \le \frac{1}{M_\delta^2} Var^*\left(\sqrt{l}T^*\right) = \left(\frac{1}{M_\delta^2}\right) O_P(1),$$

Also, by the definition of $O_P(1)$, for δ, there exists a $M_\delta^{'} < \infty$ such that

$$\lim_{n\to\infty} P\left(\mid O_P(1) \mid > M_\delta^{'}\right) = 0.$$

If we take $M_\delta = \sqrt{\frac{M_\delta^{'}}{\delta}}$, i.e. $M_\delta^2 = \frac{M_\delta^{'}}{\delta}$, then,

$$\lim_{n\to\infty} P\left(\mid \frac{1}{M_\delta^2} O_P(1) \mid > \delta\right) = \lim_{n\to\infty} P\left(\frac{\delta}{M_\delta^{'}} \mid O_P(1) \mid > \delta\right) = \lim_{n\to\infty} P\left(\mid O_P(1) \mid > M_\delta^{'}\right) = 0.$$

Then, it follows that $P^*\left(\mid \sqrt{l}T^* \mid > M_\delta\right) = o_P(1)$, i.e. $T^* = O_{P^*}\left(1/\sqrt{l}\right)$. Therefore it follows that $V^{*(g)'}P_Z\epsilon^*/l - \sigma_{V\epsilon}^{b(g)} = O_{P^*}\left(1/\sqrt{l}\right)$, as required. This proves part (a). Parts (b) and (c) follow from proof similar to that of part (a).

The proof for parts (d) and (e) are similar, so we will only prove (d). To proceed, note that by the properties of Expectation and Trace operator,

$$E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{r_n}\right\|^2\right) = E^*\left(trace\left(\frac{\hat{\Pi}'Z'V^*V^{*'}Z\hat{\Pi}}{r_n^2}\right)\right) = \frac{trace\left(\Sigma_{VV}^b\right)}{r_n}\left(trace\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n}\right)\right).$$

When $l/r_n \to \gamma < \infty$, by $\hat{\Pi} = (Z'Z)^{-1}Z'X$ we have

$$\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{r_n} = \frac{\Pi'Z'Z\Pi}{r_n} + \frac{V'Z\Pi}{r_n} + \frac{\Pi'Z'V}{r_n} + \frac{V'P_ZV}{r_n} = Q + \left(\frac{l}{r_n}\right)\Sigma_{VV} + O_P\left(\frac{1}{\sqrt{r_n}}\right) = O_P(1)$$

Thus, $E^*\left(\left\|\frac{V^{*'}Z\hat{\Pi}}{r_n}\right\|^2\right) = \left(\frac{1}{r_n}\right) O_P(1) O_P(1) = O_P\left(\frac{1}{r_n}\right)$ because Σ_{VV}^b is bounded in probability by Lemma .3.1. It follows that $\hat{\Pi}'Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$.

The proof for parts (d') and (e') are similar, so we will only prove (d'). Similar to part (d), we have

$$E^*\left(\left\|\frac{V^{*\prime}Z\hat{\Pi}}{l}\right\|^2\right) = \frac{trace\left(\Sigma_{VV}^b\right)}{l}\left(trace\left(\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{l}\right)\right).$$

When $l/r_n \to \infty$,

$$\begin{aligned}\frac{\hat{\Pi}'Z'Z\hat{\Pi}}{l} &= \frac{\Pi'Z'Z\Pi}{l} + \frac{V'Z\Pi}{l} + \frac{\Pi'Z'V}{l} + \frac{V'P_ZV}{l} \\ &= \Sigma_{VV} + O_P\left(\frac{r_n}{l}\right) + O_P\left(\frac{\sqrt{r_n}}{l}\right) + O_P\left(\frac{1}{\sqrt{l}}\right) = O_P(1).\end{aligned}$$

Thus, we obtain $E^*\left(\left\|\frac{V^{*\prime}Z\hat{\Pi}}{l}\right\|^2\right) = O_P\left(\frac{1}{l}\right)$, and it follows that $\hat{\Pi}'Z'V^*/l = O_{P^*}\left(1/\sqrt{l}\right)$ in this case. ■

Proof of Lemma .3.3

The proof follows closely from Lemma A11 of Hansen, Hausman, and Newey (2008) by replacing their a_i with a_i^* and $\tilde{V}$ with $\tilde{V}^*$. ■

Proof of Lemma .3.4

Let $\hat{\beta}^* = \hat{\beta}^*_{std}$. Let $\bar{\Upsilon} = [0, Z\hat{\pi}]$, $\bar{V}^* = [\epsilon^*, v^*]$, $\bar{X}^* = [y^*, X^*]$ where $\{y^*, X^*\}$ are the pseudo-data generated by the standard bootstrap DGP, so that $\bar{X}^* = (\bar{\Upsilon} + \bar{V}^*)D$ for

$$D = \begin{bmatrix} 1 & 0 \\ \hat{\beta} & 1 \end{bmatrix}$$

Note that $r_n/n \leq C$. Let $\hat{B}^* = \bar{X}^{*\prime}\bar{X}^*/n$. Then by $trace\left(\bar{\Upsilon}'\bar{\Upsilon}\right) = trace\left(\hat{\pi}'Z'Z\hat{\pi}\right)$ and $E^*\left[\bar{V}^*\bar{V}^{*\prime}\right] \leq CI_n$ in probability, we obtain

$$\begin{aligned} E^*\left[\left\|\bar{\Upsilon}'\bar{V}^*\right\|^2/n^2\right] &= trace\left(\bar{\Upsilon}'E^*\left[\bar{V}^{*\prime}\bar{V}^*\right]\bar{\Upsilon}\right)/n^2 \\ &\leq C \times trace\left(\hat{\pi}'Z'Z\hat{\pi}/n^2\right) \to 0\end{aligned}$$

in probability, so that $\bar{\Upsilon}'\bar{V}^*/n \to_{P^*} 0$. Let $\bar{\Sigma}^* = E^*\left[\bar{V}_i^*\bar{V}_i^{*\prime}\right]$. Then by standard arguments, we have $\bar{\Sigma}^* \to^P \bar{\Sigma} \equiv \begin{pmatrix} \sigma_{\epsilon\epsilon} & (1-\lambda)\,\sigma_{v\epsilon}' \\ (1-\lambda)\,\sigma_{v\epsilon} & (1-\lambda)\,\Sigma_{vv} \end{pmatrix} \geq Cdiag(I_k, 0)$ and

$$\bar{V}^{*\prime}\bar{V}^*/n - \bar{\Sigma}^* \to_{P^*} 0$$

so it follows that w.p.a.1,

$$\begin{aligned} \hat{B}^* &= D'\left(\bar{V}^{*\prime}\bar{V}^* + \bar{\Upsilon}'\bar{V}^* + \bar{V}^{*\prime}\bar{\Upsilon} + \bar{\Upsilon}'\bar{\Upsilon}\right)D/n \\ &= D'\bar{\Sigma}^*D + D'\bar{\Upsilon}'\bar{\Upsilon}D/n + o_P^*(1) \geq Cdiag(I_k, 0) \end{aligned}$$

in probability. Note that $\bar{\Upsilon}'\bar{\Upsilon}/n$ is bounded in probability, so that $\hat{B}^*$ minus a constant and it follows that

$$\begin{aligned} C &\leq (1, -\beta')\hat{B}^*(1, -\beta')' = (y^* - X^*\beta)'(y^* - X^*\beta)/n \\ &\leq C \parallel (1, -\beta') \parallel^2 = C(1 + \parallel \beta \parallel^2) \end{aligned}$$

in probability. Next, note that

$$\begin{aligned} E^*\left[\left\|\frac{\bar{\Upsilon}'\bar{V}^*}{r_n}\right\|^2\right] &\leq C \times trace\left(\bar{\Upsilon}'\bar{\Upsilon}/r_n\right)/r_n \\ &= C \times trace\left(\hat{\pi}'Z'Z\hat{\pi}\right)/r_n^2 \to 0 \end{aligned}$$

Then $\bar{\Upsilon}'\bar{V}^*/r_n \to_{P^*} 0$. Similarly, we have $\bar{\Upsilon}'P_Z\bar{V}^*/r_n \to_{P^*} 0$. Also, we have

$$\begin{aligned} \frac{1}{r_n}\left(\bar{V}^{*\prime}P_Z\bar{V}^* - \lambda_n\bar{V}^{*\prime}\bar{V}^*\right) &= \frac{1}{r_n}\left(l\bar{\Sigma}^* + O_{P^*}\left(\sqrt{l}\right) - l\bar{\Sigma}^* + O_{P^*}\left(\frac{l}{\sqrt{n}}\right)\right) \\ &= O_{P^*}\left(\frac{\sqrt{l}}{r_n} + \frac{l}{r_n\sqrt{n}}\right) \to^{P^*} 0 \end{aligned}$$

because $\sqrt{l}/r_n \to 0$.

Let $\hat{A}^* = \frac{1}{r_n}\hat{X}^{*'}P_Z\hat{X}^* - \lambda_n\hat{X}^{*'}\hat{X}^*$.

$$\frac{\bar{\Upsilon}'\bar{\Upsilon}}{r_n} = diag\left(0, \frac{\hat{\pi}'Z'Z\hat{\pi}}{r_n}\right) \geq diag(0, I_k)$$

Then by T, w.p.a.1,

$$\begin{aligned}\hat{A}^* &= (1-\lambda_n)diag\left(0, \frac{\hat{\pi}'Z'Z\hat{\pi}}{r_n}\right)\\ &\quad -\frac{1}{r_n}\left[\bar{\Upsilon}'P_Z\bar{V}^* + \bar{V}^{*'}P_Z\bar{\Upsilon} - \lambda_n\bar{V}^{*'}\bar{\Upsilon} - \lambda_n\bar{\Upsilon}'\bar{V}^* + \bar{V}^{*'}P_Z\bar{V}^* - \lambda_n\bar{V}^{*'}\bar{V}^*\right]\\ &\geq Cdiag(0, I_k)\end{aligned}$$

Note that $\sqrt{r_n}D(1, -\beta')' = \left(\sqrt{r_n}, \sqrt{r_n}(\hat{\beta} - \beta)\right)'$. It follows that w.p.a.1 by $\bar{X}_i^* = \frac{1}{\sqrt{r_n}}D'\tilde{X}_i^*$, for all β,

$$\begin{aligned}\frac{1}{r_n}(y^* - X^*\beta)'(P_Z - \lambda_n I_n)(y^* - X^*\beta) &= \frac{1}{r_n}(1, -\beta')\left[\bar{X}^{*'}P_Z\bar{X}^* - \lambda_n\bar{X}^{*'}\bar{X}^*\right](1, -\beta')'\\ &= \frac{1}{r_n}(1, -\beta')D'r_n\hat{A}^*D(1, -\beta')'\\ &\geq C \parallel \beta - \hat{\beta} \parallel^2\end{aligned}$$

Moreover, let

$$\hat{Q}^*(\beta) = \frac{\frac{1}{r_n}(y^* - X^*\beta)'(P_Z - \lambda_n I_n)(y^* - X^*\beta)}{\frac{1}{n}(y^* - X^*\beta)(y^* - X^*\beta)}$$

Note that

$$\hat{\beta}^* = argmin_\beta\hat{Q}^*(\beta)$$

Also it is easy to see that

$$\frac{1}{r_n}\epsilon^{*'}(P_Z - \lambda_n I_n)\epsilon^* \to_{P^*} 0$$

so that by $\epsilon^{*\prime}\epsilon^*/n \geq C$ w.p.a.1, $\hat{Q}^*(\hat{\beta}) \to_{P^*} 0$. Therefore,

$$0 \leq \hat{Q}^*(\hat{\beta}^*) \leq \hat{Q}^*(\hat{\beta}) \to_{P^*} 0$$

and hence $\hat{Q}^*(\hat{\beta}^*) \to_{P^*} 0$. By $(y^* - X^*\beta)'(y^* - X^*\beta)/n \leq C(1+ \| \beta \|^2)$, it follows that

$$0 \leq \frac{\| \hat{\beta}^* - \hat{\beta} \|^2}{1+ \| \hat{\beta}^* \|^2} \leq C\hat{Q}^*(\hat{\beta}^*) \to_{P^*} 0$$

Now we show that if $\| \hat{\beta}^* - \hat{\beta} \|^2 /(1+ \| \hat{\beta}^* \|^2) \to_{P^*} 0$, then $\| \hat{\beta}^* - \hat{\beta} \| \to_{P^*} 0$.

When $\| \hat{\beta}^* \| \geq a \equiv 2 \| \hat{\beta} \| + (1 + 2 \| \hat{\beta} \|^2)^{1/2}$, by subtracting $2 \| \hat{\beta} \|$ and squaring we have

$$\begin{aligned} (\| \hat{\beta}^* \| - 2 \| \hat{\beta} \|)^2 &= \| \hat{\beta}^* \|^2 - 4 \| \hat{\beta}^* \| \| \hat{\beta} \| + 4 \| \hat{\beta} \|^2 \\ &\geq 1 + 2 \| \hat{\beta} \|^2 \end{aligned}$$

Substracting $2 \| \hat{\beta} \|^2$, adding $\| \hat{\beta}^* \|^2$, and dividing by 2 gives

$$(\| \hat{\beta}^* \| - \| \hat{\beta} \|)^2 \geq (1+ \| \hat{\beta}^* \|^2)/2$$

Note that when $\| \hat{\beta}^* \| \geq a$,

$$\frac{\| \hat{\beta}^* - \hat{\beta} \|^2}{1+ \| \hat{\beta}^* \|^2} \geq \frac{\| \hat{\beta}^* - \hat{\beta} \|^2}{2(\| \hat{\beta}^* \| - \| \hat{\beta} \|)^2} \geq 1/2$$

It follows that $\| \hat{\beta}^* \| < a$ w.p.a.1, and hence $1+ \| \hat{\beta}^* \|^2 < 1 + a^2$ and

$$\| \hat{\beta}^* - \hat{\beta} \|^2 \leq (1 + a^2) \frac{\| \hat{\beta}^* - \hat{\beta} \|^2}{1+ \| \hat{\beta}^* \|^2} \to_{P^*} 0$$

■

Proof of Lemma .3.5

By Lemma .3.2, we have $\epsilon^{*'} P_Z \epsilon^* / l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$. Similarly, one can show that $\epsilon^{*'}\epsilon^*/n = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{n}\right)$. Also, by standard arugments, we have $\epsilon^{*'}\epsilon^* = O_{P^*}(n)$. Then

$$\begin{aligned} \lambda^* - \lambda_n &= \frac{l}{\epsilon^{*'}\epsilon^*}\left(\frac{\epsilon^{*'}P_Z\epsilon^*}{l} - \sigma^b_{\epsilon\epsilon} - \left(\frac{\epsilon^{*'}\epsilon^*}{n} - \sigma^b_{\epsilon\epsilon}\right)\right) \\ &= O_{P^*}\left(\frac{l}{n}\right)\left\{O_{P^*}\left(\frac{1}{\sqrt{l}}\right) + O_{P^*}\left(\frac{1}{\sqrt{n}}\right)\right\} = O_{P^*}\left(\frac{\sqrt{l}}{n}\right) \end{aligned}$$

■

Proof of Lemma .3.6

Note that by standard arguments $X^{*'}X^* = O_{P^*}(n)$ and $X^{*'}\epsilon^* = O_{P^*}(n)$. Therefore, in Case (I), $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}X^*}{r_n} = O_{P^*}\left(\frac{\delta^\lambda_n n}{r_n}\right)$ and $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}\epsilon^*}{r_n} = O_{P^*}\left(\frac{\delta^\lambda_n n}{r_n}\right)$. Also, by Lemma .3.5 and by $l/r_n \to \gamma < \infty$ in Case (I)

$$\begin{aligned} (\lambda^* - \lambda_n)\frac{X^{*'}X^*}{r_n} &= O_{P^*}\left(\frac{\sqrt{l}}{n}\cdot\frac{n}{r_n}\right) = O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ (\lambda^* - \lambda_n)\frac{X^{*'}\epsilon^*}{r_n} &= O_{P^*}\left(\frac{\sqrt{l}}{n}\cdot\frac{n}{r_n}\right) = O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \end{aligned}$$

Also, by the results in Lemma .3.2, we have $\frac{1}{r_n}\left(X^{*'}P_ZX^* - \lambda_n X^{*'}X^*\right) = \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ and $\frac{1}{r_n}\left(X^{*'}P_Z\epsilon^* - \lambda_n X^{*'}\epsilon^*\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$. Putting these results together, we obtain

$$\begin{aligned} \frac{1}{r_n}\left(X^{*'}P_ZX^* - \hat{\lambda}^*X^{*'}X^*\right) &= \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \frac{\delta^\lambda_n n}{r_n}\right) \\ \frac{1}{r_n}\left(X^{*'}P_Z\hat{\epsilon}^* - \hat{\lambda}^*X^{*'}\hat{\epsilon}^*\right) &= O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta^\beta_n + \frac{\delta^\lambda_n n}{r_n}\right) \end{aligned}$$

by the triangle inequality and by the fact that

$$\frac{1}{r_n}\left(X^{*'}P_Z\hat{\epsilon}^* - \hat{\lambda}^* X^{*'}\hat{\epsilon}^*\right) = \frac{1}{r_n}\left(X^{*'}P_Z\epsilon^* - \hat{\lambda}^* X^{*'}\epsilon^*\right) - \left(\frac{1}{r_n}(X^{*'}P_Z X^* - \hat{\lambda}^* X^{*'}X^*)\right)(\hat{\beta}^* - \hat{\beta}).$$

Using similar arguments, we obtain that for Case (II) $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}X^*}{l} = O_{P^*}\left(\frac{\delta_n^\lambda n}{l}\right)$, $(\hat{\lambda}^* - \lambda^*)\frac{X^{*'}\epsilon^*}{l} = O_{P^*}\left(\frac{\delta_n^\lambda n}{l}\right)$, $(\lambda^* - \lambda_n)\frac{X^{*'}X^*}{l} = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, and $(\lambda^* - \lambda_n)\frac{X^{*'}\epsilon^*}{l} = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$. Also, by the results in Lemma .3.2, $\frac{1}{l}\left(X^{*'}P_Z X^* - \lambda_n X^{*'}X^*\right) = \bar{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ and $\frac{1}{l}\left(X^{*'}P_Z\epsilon^* - \lambda_n X^{*'}\epsilon^*\right) = O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$. Then, the conclusion follows by the triangle inequality. ■

Proof of Lemma .3.7

Let $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*) = \hat{\epsilon}^{*'}\hat{\epsilon}^*/n$, then for Case (I),

$$\begin{aligned}
&\frac{\hat{\epsilon}^{*'}P_Z\hat{\epsilon}^*}{\hat{\epsilon}^{*'}\hat{\epsilon}^*} - \frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}\epsilon^*} = \frac{1}{\hat{\epsilon}^{*'}\hat{\epsilon}^*}\left(\hat{\epsilon}^{*'}P_Z\hat{\epsilon}^* - \epsilon^{*'}P_Z\epsilon^* - \lambda^*\left(\hat{\epsilon}^{*'}\hat{\epsilon}^* - \epsilon^{*'}\epsilon^*\right)\right)\\
&= \frac{r_n}{n\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)}\left((\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_Z X^* - \lambda^* X^{*'}X^*}{r_n}\right)(\hat{\beta}^* - \hat{\beta}) - 2(\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_Z\epsilon^* - \lambda^* X^{*'}\epsilon^*}{r_n}\right)\right)\\
&= O_P^*\left(\frac{r_n(\delta_n^\beta)^2}{n}\right)
\end{aligned}$$

because $\left(\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\right)^{-1} = O_{P^*}(1)$, $\frac{1}{r_n}\left(X^{*'}P_Z X^* - \lambda^* X^{*'}X^*\right) = O_{P^*}(1)$, $\frac{1}{r_n}\left(X^{*'}P_Z\epsilon^* - \lambda^* X^{*'}\epsilon^*\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ by Lemma .3.6 with $\hat{\lambda}^* = \lambda^*$ and $\delta_n^\lambda = \delta_n^\beta = 0$.

Using similar arguments, we obtain that for Case (II)

$$\begin{aligned}
&\frac{\hat{\epsilon}^{*'}P_Z\hat{\epsilon}^*}{\hat{\epsilon}^{*'}\hat{\epsilon}^*} - \frac{\epsilon^{*'}P_Z\epsilon^*}{\epsilon^{*'}\epsilon^*}\\
&= \frac{l}{n\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)}\left((\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_Z X^* - \breve{\lambda}^* X^{*'}X^*}{l}\right)(\hat{\beta}^* - \hat{\beta}) - 2(\hat{\beta}^* - \hat{\beta})'\left(\frac{X^{*'}P_Z\epsilon^* - \breve{\lambda}^* X^{*'}\epsilon^*}{l}\right)\right)\\
&= O_P^*\left(\frac{l(\delta_n^\beta)^2}{n}\right)
\end{aligned}$$

■

Proof of Lemma .3.8

Let $\bar{\epsilon}^* = y^* - X^*\bar{\beta}$ and $\bar{q}^* = X^*\bar{\epsilon}^*/\bar{\epsilon}^{*\prime}\bar{\epsilon}^*$. Suppose $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and $\hat{\beta}$. Then differentiating gives

$$\begin{aligned} -\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) &= X^{*\prime}P_ZX^* - \frac{\bar{\epsilon}^{*\prime}P_Z\bar{\epsilon}^*}{\bar{\epsilon}^{*\prime}\bar{\epsilon}^*}X^{*\prime}X^* - X^{*\prime}\bar{\epsilon}^*\frac{\bar{\epsilon}^{*\prime}P_ZX^*}{\bar{\epsilon}^{*\prime}\bar{\epsilon}^*} \\ &\quad -\frac{X^{*\prime}P_Z\bar{\epsilon}^*}{\bar{\epsilon}^{*\prime}\bar{\epsilon}^*}\bar{\epsilon}^{*\prime}X^* + 2\frac{\bar{\epsilon}^{*\prime}P_Z\bar{\epsilon}^*}{(\bar{\epsilon}^{*\prime}\bar{\epsilon}^*)^2}X^{*\prime}\bar{\epsilon}^*\bar{\epsilon}^{*\prime}X^* \\ &= X^{*\prime}P_ZX^* - \bar{\lambda}^*X^{*\prime}X^* + \bar{q}^*\hat{D}^*(\bar{\beta}^*)' + \hat{D}^*(\bar{\beta}^*)\bar{q}^{*\prime} \end{aligned}$$

where $\bar{\lambda}^* = \bar{\epsilon}^{*\prime}P_Z\bar{\epsilon}^*/\bar{\epsilon}^{*\prime}\bar{\epsilon}^*$. By Lemma .3.7, we have $\bar{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{(\delta_n^\beta)^2r_n}{n}\right)$ for Case (I). Then by Lemma .3.6 with $\delta_n^\lambda = \frac{(\delta_n^\beta)^2r_n}{n}$, we obtain

$$\begin{aligned} \frac{1}{r_n}\left(X^{*\prime}P_ZX^* - \bar{\lambda}^*X^{*\prime}X^*\right) &= \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} + (\delta_n^\beta)^2\right) \\ \frac{1}{r_n}\hat{D}^*(\bar{\beta}^*) &= O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta_n^\beta\right) \end{aligned}$$

Note that by standard argument $\bar{q}^* = O_{P^*}(1)$, hence $\frac{1}{r_n}\left(\hat{D}^*(\bar{\beta}^*)\bar{q}^*\right) = \frac{1}{r_n}\hat{D}^*(\bar{\beta}^*)O_{P^*}(1) = O_{P^*}\left(\frac{1}{\sqrt{r_n}} + \delta_n^\beta\right)$. The conclusion then follows by the triangle inequality. For Case (II), note that by similar arguments, one can obtain $\frac{1}{l}\left(X^{*\prime}P_ZX^* - \bar{\lambda}^*X^{*\prime}X^*\right) = \bar{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}} + (\delta_n^\beta)^2\right)$, $\frac{1}{l}\hat{D}^*(\bar{\beta}^*) = O_{P^*}\left(\frac{1}{\sqrt{l}} + \delta_n^\beta\right)$, and $\frac{1}{l}\left(\hat{D}^*(\bar{\beta}^*)\bar{q}^*\right) = O_{P^*}\left(\frac{1}{\sqrt{l}} + \delta_n^\beta\right)$. ■

Proof of Lemma .3.9

Note that by Lemma .3.5, $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$. Also, by standard argument, $\tilde{V}^{*\prime}\epsilon^* = O_{P^*}(n)$. Moreover, For Case (I) we have $\hat{\Pi}'Z'\epsilon^*/\sqrt{r_n} =$

$O_{P^*}\left(1/\sqrt{r_n}\right)$, thus

$$\begin{aligned}
\frac{1}{\sqrt{r_n}}\hat{D}^*(\hat{\beta}) &= \frac{1}{\sqrt{r_n}}\left(X^{*'}P_Z\varepsilon^* - \lambda^* X^{*'}\varepsilon^*\right) \\
&= \frac{1}{\sqrt{r_n}}\left\{\left(X^* - \epsilon^*\left(\frac{\sigma^{b'}_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}}\right)\right)' P_Z\epsilon^* - \lambda^*\left(X^* - \epsilon^*\left(\frac{\sigma^{b'}_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}}\right)\right)'\epsilon^*\right\} \\
&= \frac{1}{\sqrt{r_n}}\left\{\hat{\Pi}'Z'\epsilon^* + \tilde{V}^{*'}P_Z\epsilon^* - \left(Z\hat{\Pi} + \tilde{V}^*\right)'\epsilon^*\left[\lambda_n + O^*_P\left(\frac{\sqrt{l}}{n}\right)\right]\right\} \\
&= \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)
\end{aligned}$$

where the last equality follows by noting that

$$\frac{1}{\sqrt{r_n}}\left(Z\hat{\Pi} + \tilde{V}^*\right)'\epsilon^* O_{P^*}\left(\frac{\sqrt{l}}{n}\right) = O_{P^*}\left(\sqrt{\frac{n}{r_n}}\right)O_{P^*}\left(\frac{\sqrt{l}}{n}\right) = O_{P^*}\left(\sqrt{\frac{\lambda_n}{r_n}}\right)$$

and $\lambda_n \to \lambda \in [0,1)$ as $n \to \infty$. Using similar arguments, we obtain for Case (II)

$$\frac{1}{\sqrt{l}}\hat{D}^*(\hat{\beta}) = \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\hat{\Pi} + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$$

■

.3.2 Proofs of results for the RE Bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix B. Let $\hat{\beta}^* = \hat{\beta}^*_{re}$ throughout Appendix B. Also, P^* denotes the probability measure induced by the RE bootstrap procedure and E^* denotes the expectation under P^*.

Lemma .3.10. *Suppose that Assumptions 1-2 hold, then under $H_0 : \beta = \beta_0$, (a) $E^*\left(\epsilon_i^{*8}\right)$ and (b) $E^*\left(\|V_i^*\|^8\right)$ are bounded in probability.*

Lemma .3.11. *Suppose that Assumptions 1-2 hold, then under $H_0 : \beta = \beta_0$, the following statements are true as $n \to \infty$*

$$(a)\ \ V^{*\prime} P_Z \epsilon^* / l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right);$$

$$(b)\ \ V^{*\prime} P_Z V^* / l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right);$$

$$(c)\ \ \epsilon^{*\prime} P_Z \epsilon^* / l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right),$$

in probability, in Case (I) ($l/r_n \to \gamma < \infty$),

$$(d)\ \ \tilde{\Pi}'(\beta_0) Z' V^* / r_n = O_{P^*}\left(1/\sqrt{r_n}\right);$$

$$(e)\ \ \tilde{\Pi}'(\beta_0) Z' \epsilon^* / r_n = O_{P^*}\left(1/\sqrt{r_n}\right),$$

in probability, and in Case (II) ($l/r_n \to \infty$),

$$(d')\ \ \tilde{\Pi}'(\beta_0) Z' V^* / l = O_{P^*}\left(1/\sqrt{l}\right);$$

$$(e')\ \ \tilde{\Pi}'(\beta_0) Z' \epsilon^* / l = O_{P^*}\left(1/\sqrt{l}\right),$$

in probability, where $\sigma^b_{V\epsilon} \equiv E^*\left(V_i^* \epsilon_i^*\right)$, $\Sigma^b_{VV} \equiv E^*\left(V_i^* V_i^{*\prime}\right)$ and $\sigma^b_{\epsilon\epsilon} \equiv E^*\left(\epsilon_i^{*2}\right)$.

To proceed, let $A^* = diag(a_1^*, ..., a_n^*)$ where $a_i^* = \epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}, i = 1, .., n$.

Lemma .3.12. *Suppose that Assumptions 1-2 hold, then both in Case (I) and in Case (II),*

$$(a)\ \ \tilde{V}^{*\prime} A^* \tilde{V}^* / n = E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*\prime}\right) + O_{P^*}\left(1/\sqrt{n}\right);$$

$$(b)\ \ \tilde{V}^{*\prime} P_Z A^* \tilde{V}^* / n = \lambda_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*\prime}\right) + O_{P^*}\left(1/\sqrt{n}\right);$$

$$(c)\ \ \tilde{V}^{*\prime} D_Z A^* D_Z \tilde{V}^* / n = \lambda_n \phi_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*\prime}\right) + O_{P^*}\left(1/\sqrt{n}\right);$$

$$(d)\ \ \tilde{V}^{*\prime} P_Z A^* P_Z \tilde{V}^* / n = \lambda_n \phi_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*\prime}\right) + O_{P^*}\left(1/\sqrt{n}\right),$$

in probability.

Lemma .3.13. *Suppose that Assumptions 1-2 hold, then under $H_0 : \beta = \beta_0$, both in Case (I) and in Case (II),*

$$\hat{\beta}^* - \beta_0 = o_{P^*}(1), \textit{ in probability}$$

Let $\hat{\lambda}^* = \frac{\hat{\epsilon}^{*\prime} P_Z \hat{\epsilon}^*}{\hat{\epsilon}^{*\prime}\hat{\epsilon}^*}$, $\breve{\lambda}^* = \frac{\epsilon^{*\prime} P_Z \epsilon^*}{\epsilon^{*\prime}\epsilon^*}$ where $\hat{\epsilon}^* = y^* - X^*\hat{\beta}^*$ and $\{y^*, X^*\}$ denotes the pseudo-sample generated by the RE bootstrap.

Lemma .3.14. *Suppose that Assumptions 1-2 hold, then under $H_0 : \beta = \beta_0, \lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$.*

Lemma .3.15. *If $\hat{\lambda}^* = \lambda^* + O_{P^*}(\delta_n^\lambda)$ for $\delta_n^\lambda \to 0$, and $\hat{\beta}^* - \beta_0 = O_{P^*}(\delta_n^\beta)$ for $\delta_n^\beta \to 0$, then under $H_0 : \beta = \beta_0$, for Case (I),*

(a) $\frac{1}{r_n}\left(X^{*\prime} P_Z X^* - \hat{\lambda}^* X^{*\prime} X^*\right) = \tilde{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\lambda n/r_n\right);$

(b) $\frac{1}{r_n}\left(X^{*\prime} P_Z \hat{\epsilon}^* - \hat{\lambda}^* X^{*\prime} \hat{\epsilon}^*\right) = O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\beta + \delta_n^\lambda n/r_n\right),$

where $\tilde{H}_{I,n} = (1-\lambda_n)\left(\tilde{\Pi}'(\beta_0) Z' Z \tilde{\Pi}(\beta_0)/r_n\right)$; for Case (II),

(a′) $\frac{1}{l}\left(X^{*\prime} P_Z X^* - \hat{\lambda}^* X^{*\prime} X^*\right) = \tilde{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^\lambda n/l\right);$

(b′) $\frac{1}{l}\left(X^{*\prime} P_Z \hat{\epsilon}^* - \hat{\lambda}^* X^{*\prime} \hat{\epsilon}^*\right) = O_{P^*}\left(1/\sqrt{l} + \delta_n^\beta + \delta_n^\lambda n/l\right),$

where $\tilde{H}_{II,n} = (1-\lambda_n)\left(\tilde{\Pi}'(\beta_0) Z' Z \tilde{\Pi}(\beta_0)/l\right)$.

Lemma .3.16. *Suppose that Assumptions 1-2 hold. Suppose $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^\beta\right)$ for $\delta_n^\beta \to 0$, then under $H_0 : \beta = \beta_0$, in Case (I) $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{r_n}{n}(\delta_n^\beta)^2\right)$; in Case (II), $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(\frac{l}{n}(\delta_n^\beta)^2\right)$.*

Lemma .3.17. *Suppose that Assumptions 1-2 hold. Suppose $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^\beta\right)$ for $\delta_n^\beta \to 0$, then under $H_0 : \beta = \beta_0$, for Case (I),*

$$-\frac{1}{r_n}\left(\partial \hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = \tilde{H}_{I,n} + O_{P^*}\left(1/\sqrt{r_n} + \delta_n^\beta\right),$$

for Case (II),

$$-\frac{1}{l}\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = \tilde{H}_{II,n} + O_{P^*}\left(1/\sqrt{l} + \delta_n^\beta\right).$$

where $\bar{\beta}^*$ *lies between* β_0 *and* $\hat{\beta}^*$*.*

Lemma .3.18. *Suppose that Assumptions 1-2 hold, then under* $H_0 : \beta = \beta_0$*, in Case (I)*

$$\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) = \frac{1}{\sqrt{r_n}}\left((1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(1/\sqrt{r_n}\right)$$

and in Case (II)

$$\frac{1}{\sqrt{l}}\hat{D}^*(\beta_0) = \frac{1}{\sqrt{l}}\left((1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right)'\epsilon^* + O_{P^*}\left(1/\sqrt{l}\right)$$

where $\tilde{V} = V^* - \epsilon^*\left(\frac{\sigma^{b\prime}_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}}\right)$*.*

Proof of Theorem 3.3.2

The proof is similar to that of Theorem 3.3.1. One can show that similar to the standard bootstrap, for Case (I)

$$\sqrt{r_n}(\hat{\beta}^*-\beta_0) = -\sqrt{r_n}\left(\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\hat{D}^*(\beta_0) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0)$$

and for Case (II), we let

$$\sqrt{l}(\hat{\beta}^*-\beta_0) = -\sqrt{l}\left(\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\hat{D}^*(\beta_0) = -\left(\frac{1}{l}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{l}}\hat{D}^*(\beta_0).$$

where $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and β_0.

To proceed for Case (I), we let $W^*_{I,i} = \begin{pmatrix} \frac{1}{\sqrt{r_n}}(1-\lambda)\tilde{\Pi}'(\beta_0)Z_i\varepsilon^*_i \\ \frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}^*_i\varepsilon^*_i \end{pmatrix}$. By using arguments similar to those in the proof of Theorem 3.3.1 and by replacing $\hat{\Pi}$ with $\tilde{\Pi}(\beta_0)$, we can show that under H_0

$$\begin{cases} \sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{r_n}}(1-\lambda_n)\tilde{\Pi}'(\beta_0)Z_i\varepsilon_i^*\right\|^4\right) \to^P 0 \\ \sum_{i=1}^n E^*\left(\left\|\frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*\right\|^4\right) \to^P 0 \end{cases}$$

and

$$\sum_{i=1}^n E^*\left(W_{I,i}^* W_{I,i}^{*'}\right) \to^P \begin{pmatrix} (1-\lambda)\sigma_{\epsilon\epsilon}\tilde{H}_I & (1-\lambda)\tilde{A}' \\ (1-\lambda)\tilde{A} & (\phi-\lambda)\left(\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}}+\tilde{B}\right) \end{pmatrix} \equiv \tilde{\Psi}_I.$$

where $\tilde{H}_I = H_I + (1-\lambda)\gamma\Sigma_{\tilde{V}\tilde{V}}$, $\tilde{A} = \sqrt{1-\lambda}A$, and $\tilde{B} = \frac{1-2\lambda+\lambda\phi}{1-\lambda}B$. These results follow by the fact that under the RE bootstrap DGP and under H_0,

$$E^*\left(\epsilon_i^{*2}\right) = \frac{n}{n-k}\left\{\frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n} - \left(\frac{1}{n}\sum_{i=1}^n \epsilon_i(\beta_0)\right)^2\right\} = \frac{\epsilon'(\beta_0)\epsilon(\beta_0)}{n} + o_P(1) \to^P \sigma_{\epsilon\epsilon};$$

$$E^*\left(V_i^*\epsilon_i^*\right) = \sqrt{\frac{n}{n-k}}\sqrt{\frac{n}{n-l}}\left\{\frac{\tilde{V}'(\beta_0)\epsilon(\beta_0)}{n} - \left(\frac{1}{n}\sum_{i=1}^n \epsilon_i(\beta_0)\right)\left(\frac{1}{n}\sum_{i=1}^n \tilde{V}_i(\beta_0)\right)\right\}$$

$$= \sqrt{\frac{n}{n-l}}\frac{\tilde{V}'(\beta_0)\epsilon(\beta_0)}{n} + o_P(1)$$

$$= \sqrt{\frac{n}{n-l}}\left(\frac{X'M_Z\epsilon(\beta_0)}{n} + \frac{X'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\frac{\epsilon'(\beta_0)P_Z\epsilon(\beta_0)}{n}\right) + o_P(1) \to^P \frac{\sigma_{V\epsilon}}{\sqrt{1-\lambda}};$$

$$E^*\left(V_i^*V_i^{*'}\right) = \frac{n}{n-l}\left\{\frac{\tilde{V}'(\beta_0)\tilde{V}(\beta_0)}{n} - \left(\frac{1}{n}\sum_{i=1}^n \tilde{V}_i(\beta_0)\right)^2\right\}$$

$$= \frac{n}{n-l}\frac{\tilde{V}'(\beta_0)\tilde{V}(\beta_0)}{n} + o_P(1)$$

$$= \frac{n}{n-l}\left\{\frac{X'M_ZX}{n} + \frac{X'M_Z\epsilon(\beta_0)}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\frac{\epsilon'(\beta_0)P_Z\epsilon(\beta_0)}{n}\frac{\epsilon'(\beta_0)M_ZX}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right\} + o_P(1)$$

$$\to^P \Sigma_{VV} + \frac{\lambda}{1-\lambda}\frac{\sigma_{V\epsilon}\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}},$$

Notice that under H_0, $\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0)\to^P 0$ since

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0) &= \frac{1}{n}\sum_{i=1}^{n}\left(X-Z\tilde{\Pi}(\beta_0)\right)_i=\frac{1}{n}\sum_{i=1}^{n}\left(X-Z(Z'Z)^{-1}Z'\left[X-\epsilon(\beta_0)\frac{\epsilon'(\beta_0)M_ZX}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\right]\right)_i \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i+\left(\frac{1}{n}\sum_{i=1}^{n}\left(P_Z\epsilon(\beta_0)\right)_i\right)\frac{\epsilon'(\beta_0)M_ZX}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\left(P_Z\epsilon(\beta_0)\right)_i\right)\frac{\epsilon'(\beta_0)M_ZX}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)},
\end{aligned}$$

where $\frac{1}{n}\sum_{i=1}^{n}\hat{V}_i=0$ given that the Z contains a column of ones. Also, we have $\frac{\epsilon'(\beta_0)M_ZX}{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}\to^P\frac{\sigma'_{V\epsilon}}{\sigma_{\epsilon\epsilon}}$ under H_0. Therefore, to conclude that $\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0)\to^P 0$, we only need to show that $\frac{1}{n}\sum_{i=1}^{n}\left(P_Z\epsilon(\beta_0)\right)_i\to^P 0$ under H_0. Notice that under H_0, $E\left(\frac{1}{n}\sum_{i=1}^{n}\left(P_Z\epsilon(\beta_0)\right)_i\right)=0$ and

$$\begin{aligned}
Var\left(\frac{1}{n}\sum_{i=1}^{n}\left(P_Z\epsilon(\beta_0)\right)_i\right) &= \frac{1}{n^2}\sum_{i,j=1}^{n}E\left(\left(P_Z\epsilon(\beta_0)\right)_i\left(P_Z\epsilon(\beta_0)\right)_j\right) \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}P_{ik}P_{jl}E\left(\epsilon_k\epsilon_l\right) \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n}\sum_{k=1}^{n}P_{ik}P_{jk}E\left(\epsilon_k^2\right)=\frac{\sigma_{\epsilon\epsilon}}{n^2}\sum_{i,j=1}^{n}\sum_{k=1}^{n}P_{ik}P_{jk} \\
&= \frac{\sigma_{\epsilon\epsilon}}{n^2}\sum_{i,j=1}^{n}\sum_{k=1}^{n}P_{ik}P_{kj}=\frac{\sigma_{\epsilon\epsilon}}{n^2}\sum_{i,j=1}^{n}P_{ij}\to 0,
\end{aligned}$$

since by the Frobenius norm,

$$\left|\sum_{i,j=1}^{n}P_{ij}\right|\le\sum_{i,j=1}^{n}|P_{ij}|\equiv\parallel P_Z\parallel_1\le\sqrt{n}\parallel P_Z\parallel_2=\sqrt{n}\parallel P_Z\parallel_F=\sqrt{nl}.$$

This concludes the proof of $\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_i(\beta_0)\to^P 0$.

Putting the result of $E^*\left(\epsilon_i^{*2}\right)$, $E^*\left(V_i^*\epsilon_i^*\right)$ and $E^*\left(V_i^*V_i^{*'}\right)$ together, we obtain that $E^*\left(\tilde{V}_i^*\tilde{V}_i^{*'}\right)\to^P\Sigma_{\tilde{V}\tilde{V}}$, i.e., the RE bootstrap consistently estimates

$\Sigma_{\tilde{V}\tilde{V}}$. Similarly, we can show that

$$E^*\left(\epsilon_i^{*2}\tilde{V}_i^{*'}\right) = E^*\left\{\epsilon_i^{*2}\left(V_i^{*'} - \epsilon_i^*\frac{\sigma_{V\epsilon}^{b'}}{\sigma_{\epsilon\epsilon}^b}\right)\right\}$$
$$= \sqrt{\frac{n}{n-l}}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i^2(\beta_0)\tilde{V}_i(\beta_0)\right) - \frac{1}{n}\sum_{i=1}^n \epsilon_i^3(\beta_0)\left(\frac{\sigma_{V\epsilon}^{b'}}{\sigma_{\epsilon\epsilon}^b}\right) + o_P(1).$$

Let $\tilde{a} = (\epsilon_1^2(\beta_0) - \sigma_{\epsilon\epsilon}, ..., \epsilon_n^2(\beta_0) - \sigma_{\epsilon\epsilon})$ and $a = (\epsilon_1^2 - \sigma_{\epsilon\epsilon}, ..., \epsilon_n^2 - \sigma_{\epsilon\epsilon})$, then we have

$$\frac{1}{n}\sum_{i=1}^n \epsilon_i^2(\beta_0)\tilde{V}_i(\beta_0) = \frac{(\tilde{a}-a)'\tilde{V}(\beta_0)}{n} + \frac{a'\tilde{V}(\beta_0)}{n} = o_P(1) + \frac{a'\tilde{V}(\beta_0)}{n}.$$

Also notice that under H_0,

$$\frac{a'\tilde{V}(\beta_0)}{n} = \frac{a'\left(V + Z\left(\Pi - \tilde{\Pi}(\beta_0)\right)\right)}{n} \to^P (1-\lambda)E\left(\epsilon_i^2 V_i\right) + \lambda E\left(\epsilon_i^3\right)\left(\frac{\sigma_{V\epsilon}'}{\sigma_{\epsilon\epsilon}}\right).$$

Therefore, we obtain under H_0, $E^*\left(\epsilon_i^{*2}\tilde{V}_i^{*'}\right) \to^P \sqrt{1-\lambda}E\left(\epsilon_i^2\tilde{V}_i'\right)$. Finally, using Lemma A10 and A11 in Hansen, Hausman, and Newey (2008) and using arguments similar to the case of standard bootstrap, we obtain

$$E^*\left\{\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right)\tilde{V}_i^*\tilde{V}_i^{*'}\right\} \to^P \frac{1-2\lambda+\lambda\phi}{1-\lambda}E\left((\epsilon_i^2 - \sigma_{\epsilon\epsilon})\tilde{V}_i\tilde{V}_i'\right)$$

Then using similar arguments as in Appendix A, we obtain that for Case (I) and under H_0

$$\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) = \frac{1}{\sqrt{r_n}}\left\{(1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \to^{d^*} N\left(0, \tilde{\Upsilon}_I\right),$$

with $\tilde{\Upsilon}_I = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{\tilde{H}_I + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\} + (1-\lambda)\sqrt{\gamma}\left\{\tilde{A}+\tilde{A}'\right\} + \gamma\tilde{B}$ and then

$$\sqrt{r_n}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{r_n}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1}\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) \rightarrow^{d^*} N\left(0, \tilde{\Lambda}_I\right)$$

where $\tilde{\Lambda}_I = \tilde{H}_I^{-1}\tilde{\Upsilon}_I\tilde{H}_I^{-1}$.

For Case (II) we let $W^*_{II,i} = \left(\frac{1}{\sqrt{l}}(1-\lambda_n)\tilde{\Pi}'(\beta_0)Z_i\varepsilon_i^*, \frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_i^*\varepsilon_i^*\right)'$. Then similar to the standard bootstrap, we have

$$\sum_{i=1}^{n} E^*\left(W^*_{II,i}W^{*\prime}_{II,i}\right) \rightarrow_P \begin{pmatrix} (1-\lambda)\sigma_{\epsilon\epsilon}\tilde{H}_{II} & 0 \\ 0 & (\phi-\lambda)\left(\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}} + \tilde{B}\right) \end{pmatrix} \equiv \tilde{\Psi}_{II},$$

under H_0; where $\tilde{H}_{II} = (1-\lambda)\Sigma_{\tilde{V}\tilde{V}}$. Then, similar to Case (I), we obtain

$$\frac{1}{\sqrt{l}}\hat{D}^*(\beta_0) = \frac{1}{\sqrt{l}}\left\{(1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{l}}\right) \rightarrow^{d^*} N\left(0, \tilde{\Upsilon}_{II}\right),$$

with $\tilde{\Upsilon}_{II} = \sigma_{\epsilon\epsilon}\left\{(1-\lambda)\tilde{H}_{II} + (1-\phi)\Sigma_{\tilde{V}\tilde{V}}\right\}$. Together with Lemma .3.17, we have

$$\sqrt{l}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{l}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1}\frac{1}{\sqrt{l}}\hat{D}^*(\beta_0) \rightarrow^{d^*} N\left(0, \tilde{\Lambda}_{II}\right),$$

where $\tilde{\Lambda}_{II} = \tilde{H}_{II}^{-1}\tilde{\Upsilon}_{II}\tilde{H}_{II}^{-1}$. ■

Proof of Corollary 3.3.2

By Theorem 3.3.2, we have when $l/r_n \rightarrow 0$, $\sqrt{r_n}(\hat{\beta}^*_{re} - \beta_0) \rightarrow^{d^*} N(0, \sigma_{\epsilon\epsilon}Q^{-1})$, which is the same as the limiting distribution of $\sqrt{r_n}(\hat{\beta} - \beta_0)$. The result then follows by Polya's Theorem. ■

Now we give the proofs for the Lemmas.

Proof of Lemma .3.10

It is similar to the proof of Lemma .3.1.

Proof of Lemma .3.11

It is similar to the proof of Lemma .3.2 by replacing $\hat{\Pi}$ with $\tilde{\Pi}(\beta_0)$.

Proof of Lemma .3.12

It is similar to the proof of Lemma .3.3.

Proof of Lemma .3.13

It is similar to the proof of Lemma .3.4.

Proof of Lemma .3.14

It is similar to the proof of Lemma .3.5.

Proof of Lemma .3.15

Note that for the RE bootstrap, by .3.11 $\frac{1}{r_n}\left(X^{*'}P_Z X^* - \lambda_n X^{*'}X^*\right) = \tilde{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ for Case (I), and $\frac{1}{l}\left(X^{*'}P_Z X^* - \lambda_n X^{*'}X^*\right) = \tilde{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$ for Case (II), then the results follows by applying same arguments as in the proof of Lemma .3.6. ■

Proof of Lemma .3.16

It is similar to the proof of Lemma .3.7. ■

Proof of Lemma .3.17

The result follows by using similar arguments as in the proof of Lemma .3.8 and replacing $\bar{H}_{I,n}$, $\bar{H}_{II,n}$ therein with $\tilde{H}_{I,n}$ and $\tilde{H}_{II,n}$. ■

Proof of Lemma .3.18

Note that by Lemma .3.14, $\lambda^* = \lambda_n + O_P^*\left(\frac{\sqrt{l}}{n}\right)$. Moreover, for Case (I) we have $\frac{\tilde{\Pi}'(\beta_0)Z'\epsilon^*}{\sqrt{r_n}} = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$ and $\frac{\tilde{V}^{*'}\epsilon^*}{\sqrt{r_n}} = O_{P^*}\left(\sqrt{\frac{n}{r_n}}\right)$ by Markov inequality.

Therefore, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) \\
= &\frac{1}{\sqrt{r_n}}\left\{\tilde{\Pi}'(\beta_0)Z'\epsilon^* + \tilde{V}^{*'}P_Z\epsilon^* - \left(Z\tilde{\Pi}(\beta_0)+\tilde{V}^*\right)'\epsilon^*\left[\lambda_n + O_{P^*}\left(\frac{\sqrt{l}}{n}\right)\right]\right\} \\
= &\frac{1}{\sqrt{r_n}}\left\{(1-\lambda_n)Z\tilde{\Pi}(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)
\end{aligned}
$$

The result for Case (II) follows from similar arguments. ■

.3.3 Proofs of results for the Modified RE Bootstrap

All the proofs of the Lemmas are relegated at the end of Appendix B. Let $\hat{\beta}^* = \hat{\beta}^*_m$ throughout Appendix C. Also, P^* denotes the probability measure induced by the MRE bootstrap procedure and E^* denotes the expectation under P^*.

Lemma .3.19. *Suppose that Assumptions 1-2 hold, then under $H_0: \beta = \beta_0$, (a) $E^*\left(\epsilon_i^{*8}\right)$ and (b) $E^*\left(\|V_i^*\|^8\right)$ are bounded in probability.*

Lemma .3.20. *Suppose that Assumptions 1-2 hold, then under $H_0: \beta = \beta_0$, the following statements are true as $n \to \infty$*

(a) $V^{*'}P_Z\epsilon^*/l = \sigma^b_{V\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$;

(b) $V^{*'}P_ZV^*/l = \Sigma^b_{VV} + O_{P^*}\left(1/\sqrt{l}\right)$;

(c) $\epsilon^{*'}P_Z\epsilon^*/l = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$,

in probability, and both in Case (I) and Case (II)

(d) $\tilde{\Pi}'_m(\beta_0)Z'V^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$;

(e) $\tilde{\Pi}'_m(\beta_0)Z'\epsilon^*/r_n = O_{P^*}\left(1/\sqrt{r_n}\right)$, *in probability, where* $\sigma^b_{V\epsilon} \equiv E^*(V_i^*\epsilon_i^*)$, $\Sigma^b_{VV} \equiv E^*\left(V_i^* V_i^{*'}\right)$ *and* $\sigma^b_{\epsilon\epsilon} \equiv E^*(\epsilon_i^{*2})$.

To proceed, let $A^* = diag(a_1^*, ..., a_n^*)$ where $a_i^* = \epsilon_i^{*2} - \sigma^b_{\epsilon\epsilon}, i = 1, .., n.$

Lemma .3.21. *Suppose that Assumptions 1-2 hold, then both in Case (I) and in Case (II),*

(a) $\tilde{V}^{*'} A^* \tilde{V}^*/n = E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*'}\right) + O_{P^*}(1/\sqrt{n});$

(b) $\tilde{V}^{*'} P_Z A^* \tilde{V}^*/n = \lambda_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*'}\right) + O_{P^*}(1/\sqrt{n});$

(c) $\tilde{V}^{*'} D_Z A^* D_Z \tilde{V}^*/n = \lambda_n \phi_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*'}\right) + O_{P^*}(1/\sqrt{n});$

(d) $\tilde{V}^{*'} P_Z A^* P_Z \tilde{V}^*/n = \lambda_n \phi_n E^*\left(a_i^* \tilde{V}_i^* \tilde{V}_i^{*'}\right) + O_{P^*}(1/\sqrt{n}),$

in probability.

Lemma .3.22. *Suppose that Assumptions 1-2 hold, then both in Case (I) and in Case (II),*

$$\hat{\beta}^* - \beta_0 = o_{P^*}(1)$$

Let $\hat{\lambda}^* = \frac{\hat{\epsilon}^{*'} P_Z \hat{\epsilon}^*}{\hat{\epsilon}^{*'} \hat{\epsilon}^*}$, $\lambda^* = \frac{\epsilon^{*'} P_Z \epsilon^*}{\epsilon^{*'} \epsilon^*}$ where $\hat{\epsilon}^* = y^* - X^* \hat{\beta}^*$ and $\{y^*, X^*\}$ denotes the pseudo-sample generated by the MRE bootstrap.

Lemma .3.23. *Suppose that Assumptions 1-2 hold, then* $\lambda^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$.

Lemma .3.24. *If* $\hat{\lambda}^* = \lambda^* + O_{P^*}(\delta_n^\lambda)$ *for* $\delta_n^\lambda \to 0$, *and* $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n^\beta \to 0$, *then for both Case (I) and Case (II)*

$$\begin{aligned} \frac{1}{r_n}\left(X^{*'} P_Z X^* - \hat{\lambda}^* X^{*'} X^*\right) &= \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n + \delta_n^\lambda n/r_n\right) \\ \frac{1}{r_n}\left(X^{*'} P_Z \hat{\epsilon}^* - \hat{\lambda}^* X^{*'} \hat{\epsilon}^*\right) &= O_{P^*}\left(\sqrt{l}/r_n + \delta_n^\beta + \delta_n^\lambda n/r_n\right) \end{aligned}$$

where $\tilde{H}_{m,n} = (1 - \lambda_n)\left(\tilde{\Pi}_m'(\beta_0) Z' Z \tilde{\Pi}_m(\beta_0)/r_n\right)$.

Lemma .3.25. *Suppose that Assumptions 1-2 hold. Suppose* $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n^\beta \to 0$*, then under* H_0*, both in Case (I) and Case (II)* $\hat{\lambda}^* = \lambda^* + O_{P^*}\left(r_n(\delta_n^\beta)^2/n\right)$.

Lemma .3.26. *Suppose that Assumptions 1-2 hold. Suppose* $\hat{\beta}^* - \beta_0 = O_{P^*}\left(\delta_n^\beta\right)$ *for* $\delta_n \to 0$ *then under* H_0*, for both Case (I) and Case (II)*

$$-\frac{1}{r_n}\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = \tilde{H}_{m,n} + O_{P^*}\left(\sqrt{l}/r_n + \delta_n^\beta\right)$$

where $\bar{\beta}^*$ *lies between* β_0 *and* $\hat{\beta}^*$.

Lemma .3.27. *Suppose that Assumptions 1-2 hold, then under* H_0*, for Case (I)*

$$\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) = \frac{1}{\sqrt{r_n}}\left\{(1-\lambda_n)Z\tilde{\Pi}_m(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + O_{P^*}\left(1/\sqrt{r_n}\right),$$

and in Case (II)

$$\frac{r_n}{\sqrt{l}}\hat{D}^*(\beta_0) = \frac{r_n}{\sqrt{l}}\left\{(1-\lambda_n)Z\tilde{\Pi}_m(\beta_0) + P_Z\tilde{V}^* - \lambda_n\tilde{V}^*\right\}'\epsilon^* + O_{P^*}\left(\sqrt{l}/r_n\right).$$

Proof of Theorem 3.3.3

Let $\bar{\beta}^*$ be the mean value lying between $\hat{\beta}^*$ and β_0. For Case (I) we have

$$\sqrt{r_n}(\hat{\beta}^* - \beta_0) = -\sqrt{r_n}\left(\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\hat{D}^*(\beta_0) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0)$$

and for Case (II), we have

$$\frac{r_n}{\sqrt{l}}(\hat{\beta}^* - \beta_0) = -\frac{r_n}{\sqrt{l}}\left(\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\hat{D}^*(\beta_0) = -\left(\frac{1}{r_n}\frac{\partial\hat{D}^*(\bar{\beta}^*)}{\partial\beta}\right)^{-1}\frac{1}{\sqrt{l}}\hat{D}^*(\beta_0).$$

Then, we let $W_{i,m}^{*}=\begin{pmatrix}\frac{1}{\sqrt{r_n}}(1-\lambda_n)\tilde{\Pi}_m'(\beta_0)Z_i\varepsilon_{i,m}^{*}\\ \frac{1}{\sqrt{l}}(P_{ii}-\lambda_n)\tilde{V}_{i,m}^{*}\varepsilon_{i,m}^{*}\end{pmatrix}$ and we obtain,

$$\begin{aligned}
&\sum_{i=1}^{n}E^{*}\left(W_{i,m}^{*}W_{i,m}^{*\prime}\right)\\
=&\begin{pmatrix}\sum_{i=1}^{n}\frac{(1-\lambda_n)^2}{r_n}\tilde{\Pi}_m'(\beta_0)Z_iZ_i'\tilde{\Pi}_m(\beta_0)E^{*}\left(\varepsilon_{i,m}^{*2}\right) & \sum_{i=1}^{n}\frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)\tilde{\Pi}_m'(\beta_0)Z_iE^{*}\left(\varepsilon_{i,m}^{*2}\tilde{V}_{i,m}^{*\prime}\right)\\ E^{*}\left(\varepsilon_{i,m}^{*2}\tilde{V}_{i,m}^{*}\right)\sum_{i=1}^{n}\frac{1-\lambda_n}{\sqrt{lr_n}}(P_{ii}-\lambda_n)Z_i'\tilde{\Pi}_m(\beta_0) & \sum_{i=1}^{n}\frac{1}{l}(P_{ii}-\lambda_n)^2E^{*}\left(\varepsilon_{i,m}^{*2}\tilde{V}_{i,m}^{*}\tilde{V}_{i,m}^{*\prime}\right)\end{pmatrix}\\
\to_P&\begin{pmatrix}\sigma_{\epsilon\epsilon}(1-\lambda)H & 0\\ 0 & 0\end{pmatrix}\equiv\tilde{\Psi}_{m,I},
\end{aligned}$$

where the last line follows from the fact that under H_0,

$$\begin{aligned}
&\frac{\tilde{\Pi}_m'(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0)}{r_n}=\frac{\tilde{Q}(\beta_0)-l\widehat{\Sigma}_{\tilde{V}\tilde{V}}(\beta_0)}{r_n}\\
&=\frac{\Pi'Z'Z\Pi}{r_n}+O_P\left(\frac{1}{\sqrt{r_n}}\right)+\frac{l}{r_n}\left\{\Sigma_{\tilde{V}\tilde{V}}+O_P\left(\frac{1}{\sqrt{l}}\right)\right\}-\frac{l}{r_n}\left\{\Sigma_{\tilde{V}\tilde{V}}+O_P\left(\frac{1}{\sqrt{l}}\right)\right\}\\
&=\frac{\Pi'Z'Z\Pi}{r_n}+O_P\left(\frac{1}{\sqrt{r_n}}\right)+O_P\left(\frac{\sqrt{l}}{r_n}\right)\to^{P}Q
\end{aligned}$$

and from Assumption 3(a) or Assumption 3(b).

Also notice that under the MRE bootstrap DGP and under H_0

$$\begin{aligned}
E^{*}\left(\epsilon_{i,m}^{*2}\right)&=\frac{n}{n-l}\left(\frac{\epsilon'(\beta_0)M_Z\epsilon(\beta_0)}{n}\right)+o_P(1)\to_P\sigma_{\epsilon\epsilon}\\
E^{*}\left(V_{i,m}^{*}\epsilon_{i,m}^{*}\right)&=\frac{n}{n-l}\left(\frac{\hat{V}'M_Z\epsilon(\beta_0)}{n}\right)+o_P(1)\to_P\sigma_{V\epsilon}\\
E^{*}\left(V_{i,m}^{*}V_{i,m}^{*\prime}\right)&=\frac{n}{n-l}\left(\frac{\hat{V}'\hat{V}}{n}\right)+o_P(1)\to_P\Sigma_{VV}
\end{aligned}$$

which leads to the result that $E^{*}\left(\tilde{V}_{i,m}^{*}\tilde{V}_{i,m}^{*\prime}\right)\to^{P}\Sigma_{\tilde{V}\tilde{V}}$.

Proceeding as in the proof for Theorem 3.3.1 and Theorem 3.3.2, we obtain under H_0 and in case (I),

$$\sqrt{r_n}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{r_n}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1} \frac{1}{\sqrt{r_n}}\hat{D}^*(\beta_0) \to^{d^*} N\left(0, \tilde{\Lambda}_{m,I}\right),$$

where $\tilde{\Lambda}_{m,I} = H^{-1}\tilde{\Upsilon}_{m,I}H^{-1}$ and $\tilde{\Upsilon}_{m,I} = (1-\lambda)\sigma_{\epsilon\epsilon}\left\{H + \gamma\Sigma_{\tilde{V}\tilde{V}}\right\}$. Similarly, we obtain that under H_0 and in Case (II),

$$\frac{r_n}{\sqrt{l}}\left(\hat{\beta}^* - \beta_0\right) = \left(\frac{1}{r_n}\frac{\partial \hat{D}^*(\bar{\beta}^*)}{\partial \beta}\right)^{-1} \frac{1}{\sqrt{l}}\hat{D}^*(\beta_0) \to^{d^*} N\left(0, \tilde{\Lambda}_{m,II}\right),$$

where $\tilde{\Lambda}_{m,II} = H^{-1}\tilde{\Upsilon}_{m,II}H^{-1}$, $\tilde{\Upsilon}_{m,II} = (1-\lambda)\sigma_{\epsilon\epsilon}\Sigma_{\tilde{V}\tilde{V}}$. Moreover, we obtain that in Case (I), $\sqrt{r_n}(\hat{\beta} - \beta) \to^d N\left(0, \tilde{\Lambda}_{m,I}\right)$, and in Case (II), $\frac{r_n}{\sqrt{l}}(\hat{\beta} - \beta) \to^d N\left(0, \tilde{\Lambda}_{m,II}\right)$. Therefore, the bootstrap validity follows by applying Polya's Theorem to both Case (I) and Case (II). ■

Now we give the proofs for the Lemmas.

Proof of Lemma .3.19

It is similar to the proof of Lemma .3.1.

Proof of Lemma .3.20

The proof for (a)-(c) is similar to those in the proof of Lemma .3.2. The proof for (d) and (e) follows from noting that for both Case (I) and Case (II), $E^*\left\|\frac{\tilde{\Pi}'_m(\beta_0)Z'V^*}{r_n}\right\|^2 = O_P\left(\frac{1}{r_n}\right)$ and $E^*\left\|\frac{\tilde{\Pi}'_m(\beta_0)Z'\epsilon^*}{r_n}\right\|^2 = O_P\left(\frac{1}{r_n}\right)$ because $\tilde{\Pi}'_m(\beta_0)Z'Z\tilde{\Pi}_m(\beta_0) = O_P(r_n)$ for both cases. ■

Proof of Lemma .3.21

It is similar to the proof of Lemma .3.3.

Proof of Lemma .3.22

It is similar to the proof of Lemma .3.4.

Proof of Lemma .3.23

It is similar to the proof of Lemma .3.5.

Proof of Lemma .3.24

Notice that by Lemma .3.20, for both Case (I) and Case (II) $\frac{1}{r_n}\left(X^{*'}P_ZX^* - \lambda_n X^{*'}X^*\right) = \tilde{H}_{m,n} + O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right)$. Then the results follows by applying same arguments as in the proof of Lemma .3.5. ■

Proof of Lemma .3.25

Let $\hat{\sigma}^*_{\epsilon\epsilon} = \hat{\epsilon}^{*'}\hat{\epsilon}^*/n$, then in both Case (I) and Case (II), we have

$$\begin{aligned} & \hat{\epsilon}^{*'}P_Z\hat{\epsilon}^*/\hat{\epsilon}^{*'}\hat{\epsilon}^* - \epsilon^{*'}P_Z\epsilon^*/\epsilon^{*'}\epsilon^* \\ = & \frac{r_n}{n\hat{\sigma}^*_{\epsilon\epsilon}}\left\{\left(\hat{\beta}^* - \beta_0\right)'\left(\frac{X^{*'}P_ZX^* - \lambda^* X^{*'}X^*}{r_n}\right)\left(\hat{\beta}^* - \beta_0\right) - 2\left(\hat{\beta}^* - \beta_0\right)'\left(\frac{X^{*'}P_Z\epsilon^* - \lambda^* X^{*'}\epsilon^*}{r_n}\right)\right\} \\ = & O_{P^*}\left(\frac{r_n}{n}\cdot(\delta_n^\beta)^2\right) \end{aligned}$$

by $(\hat{\sigma}^*_{\epsilon\epsilon})^{-1} = O_{P^*}(1)$ and $r_n^{-1}\left(X^{*'}P_ZX^* - \lambda^* X^{*'}X^*\right) = O_{P^*}(1)$. ■

Proof of Lemma .3.26

Similar to the proof of Lemma .3.8, let $\bar{\epsilon}^* = y^* - X^*\bar{\beta}^*$ and $\bar{\gamma}^* = X^*\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$, where $\bar{\beta}^*$ lies between $\hat{\beta}^*$ and β_0. Differentiating gives

$$-\left(\partial\hat{D}^*(\bar{\beta}^*)/\partial\beta\right) = X^{*'}P_ZX^* - \bar{\lambda}^* X^{*'}X^* + \bar{\gamma}^*\hat{D}^*(\bar{\beta}^*)' + \hat{D}^*(\bar{\beta}^*)\bar{\gamma}^{*'},$$

where $\bar{\lambda}^* = \bar{\epsilon}^{*'}P_Z\bar{\epsilon}^*/\bar{\epsilon}^{*'}\bar{\epsilon}^*$. Notice that for both Case (I) and Case (II), by Lemma .3.24 and Lemma .3.25, we have

$$\begin{aligned} \frac{1}{r_n}\left(X^{*'}P_ZX^* - \bar{\lambda}^* X^{*'}X^*\right) &= \tilde{H}_{m,n} + O_{P^*}\left(\frac{\sqrt{l}}{r_n} + (\delta_n^\beta)^2\right) \\ \frac{1}{r_n}\hat{D}^*\left(\bar{\beta}^*\right) &= O_{P^*}\left(\frac{\sqrt{l}}{r_n} + \delta_n^\beta\right) \end{aligned}$$

Also, by standard argument we have $\bar{\gamma}^* = O_{P^*}(1)$, and $\frac{1}{r_n}\hat{D}^*\left(\bar{\beta}^*\right)\bar{\gamma}^* = \frac{1}{r_n}\hat{D}^*\left(\bar{\beta}^*\right)O_{P^*}(1) = O_{P^*}\left(\frac{\sqrt{l}}{r_n} + \delta_n^\beta\right)$. The conclusion then follows by triangle inequality. ■

Proof of Lemma .3.27

It is similar to the proof of Lemma .3.9.

.3.4 Proofs of results for Theorem 3.3.4

We give the proof for the case of standard bootstrap. The proof for the RE/MRE bootstraps are similar. Let $\hat{\beta}^* = \hat{\beta}^*_{std}$, $\hat{X}^*(\hat{\beta}^*) = X^* - \epsilon^*(\hat{\beta}^*)\frac{\epsilon^{*'}(\hat{\beta}^*)X^*}{\epsilon^{*'}(\hat{\beta}^*)\epsilon^*(\hat{\beta}^*)}$ and $\tilde{X}^* = X^* - \epsilon^* \frac{\sigma^{b'}_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}} = Z\hat{\Pi} + \tilde{V}^*$.

To obtain the asymptotic behavior of $\hat{\Lambda}^*(\hat{\beta}^*)$, the bootstrap analogue of the CSE, we start with the term $\hat{\Upsilon}^*(\hat{\beta}^*)$. For Case (I), notice that $\left\|\hat{\beta}^* - \hat{\beta}\right\| = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$, then

$$n^{-1}\left\|\hat{\epsilon}^*(\hat{\beta}^*) - \epsilon^*\right\|^2 \leq n^{-1}\left\|X^*\right\|^2\left\|\hat{\beta}^* - \hat{\beta}\right\|^2 \leq \left(n^{-1}\left\|X^*\right\|^2\right) O_{P^*}\left(\frac{1}{r_n}\right) = O_{P^*}\left(\frac{1}{r_n}\right)$$

by $X^{*'}X^* = O_{P^*}(n)$. It then follows by standard arguments that

$$n^{-1}\left\|\frac{X^{*'}\hat{\epsilon}^*(\hat{\beta}^*)}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - \frac{\sigma^b_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}}\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$$

and

$$\left\|\hat{X}^*(\hat{\beta}^*) - \tilde{X}^*\right\| = \left\|\hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*'}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*'}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)} - \epsilon^*\left(\frac{\sigma^{b'}_{V\epsilon}}{\sigma^b_{\epsilon\epsilon}}\right)\right\| = O_{P^*}\left(\sqrt{\frac{n}{r_n}}\right).$$

Also, $\left\|\tilde{X}^*\right\| = O_{P^*}(\sqrt{n})$. Therefore, $\left\|\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*) - \tilde{X}^{*'}\tilde{X}^*\right\| = O_{P^*}\left(\frac{n}{\sqrt{r_n}}\right)$. Then, by $\hat{\lambda}^* = \lambda_n + O_{P^*}\left(\sqrt{l}/n\right)$, we obtain

$$\left\|\hat{\lambda}^*\left(\frac{\tilde{X}^{*'}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} - \frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right)\right\| = O_{P^*}\left(\frac{l}{n}\right)O_{P^*}\left(\frac{1}{r_n}\right)O_{P^*}\left(\frac{n}{\sqrt{r_n}}\right) = O_{P^*}\left(\frac{l}{r_n\sqrt{r_n}}\right)$$

Also, by $\tilde{X}^{*'}\tilde{X}^* = O_{P^*}(n)$, we have

$$\left\|\left(\hat{\lambda}^* - \lambda_n\right)\frac{\tilde{X}^{*'}\tilde{X}^*}{r_n}\right\| = O_{P^*}\left(\frac{\sqrt{l}}{n}\right)O_{P^*}\left(\frac{n}{r_n}\right) = O_{P^*}\left(\frac{\sqrt{l}}{r_n}\right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$$

in Case (I). Putting these results together, we obtain

$$\hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} \right) = \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} \right) + \left(\hat{\lambda}^* - \lambda_n \right) \frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} + \hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} - \frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} \right)$$
$$= \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} \right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) + O_{P^*}\left(\frac{l}{r_n\sqrt{r_n}} \right) = \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} \right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right)$$

Similarly, we can show that for Case (II)

$$\hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l} \right) = \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{l} \right) + \left(\hat{\lambda}^* - \lambda_n \right) \frac{\tilde{X}^{*\prime}\tilde{X}^*}{l} + \hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l} - \frac{\tilde{X}^{*\prime}\tilde{X}^*}{l} \right)$$
$$= \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{l} \right) + O_{P^*}\left(\frac{1}{\sqrt{l}} \right)$$

Moreover, it follows from arguments similar to Lemma .3.2 that for Case (I)

$$\begin{aligned} \lambda_n \left(\frac{\tilde{V}^{*\prime}\tilde{V}^*}{r_n} \right) &= \left(\frac{l}{r_n} \right) \left(\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{n}} \right) \right) = \left(\frac{l}{r_n} \right) \Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{l}{r_n\sqrt{n}} \right) \\ &= \left(\frac{l}{r_n} \right) \Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \end{aligned}$$

where $\Sigma^b_{\tilde{V}\tilde{V}} = E^*\left(\tilde{V}_i^* \tilde{V}_i^{*\prime} \right)$, and $\lambda_n \left(\frac{\hat{\Pi}' Z' \tilde{V}^*}{r_n} \right) = O_{P^*}\left(\frac{l}{n} \right) O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) = O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right)$. Then, together with previous arguments, we obtain by $\tilde{X}^* = Z\hat{\Pi} + \tilde{V}^*$ that for Case (I)

$$\begin{aligned} \hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n} \right) &= \lambda_n \left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n} \right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \\ &= \lambda_n \left(\frac{\hat{\Pi}' Z' Z \hat{\Pi}}{r_n} \right) + \left(\frac{l}{r_n} \right) \Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{r_n}} \right) \end{aligned}$$

Similarly, we obtain for Case (II) that $\hat{\lambda}^* \left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l} \right) = \lambda_n \left(\frac{\hat{\Pi}' Z' Z \hat{\Pi}}{l} \right) +$

$\Sigma^b_{\tilde{V}\tilde{V}} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$.

Proceeding similarly for the term $\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{r_n}$, we can show that for Case (I)

$$\begin{aligned}\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{r_n} - \hat{\lambda}^*\left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right) &= \frac{\tilde{X}^{*\prime}P_Z\tilde{X}^*}{r_n} - \hat{\lambda}^*\left(\frac{\tilde{X}^{*\prime}\tilde{X}^*}{r_n}\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ &= \bar{H}_{I,n} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right),\end{aligned}$$

and for Case (II), $\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{l} - \hat{\lambda}^*\left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right) = \bar{H}_{II,n} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$.

Also, we obtain by standard arguments that $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*) = \hat{\epsilon}^{*\prime}\hat{\epsilon}^*/n = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{r_n}\right)$ for Case (I) and $\hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*) = \sigma^b_{\epsilon\epsilon} + O_{P^*}\left(1/\sqrt{l}\right)$ for Case (II). Therefore, for Case (I)

$$\begin{aligned}r_n^{-1}\hat{\Upsilon}^*_{bkk}(\hat{\beta}^*) &= \hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\left\{\left(1-2\hat{\lambda}^*(\hat{\beta}^*)\right)\left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{r_n} - \hat{\lambda}^*(\hat{\beta}^*)\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right)\right. \\ &\quad \left. +\hat{\lambda}^*(\hat{\beta}^*)\left(1-\hat{\lambda}^*(\hat{\beta}^*)\right)\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{r_n}\right\} \\ &= \sigma^b_{\epsilon\epsilon}\left\{(1-\lambda_n)\left[\bar{H}_{I,n} + \left(\frac{l}{r_n}\right)\Sigma^b_{\tilde{V}\tilde{V}}\right]\right\} + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)\end{aligned}$$

and for Case (II)

$$\begin{aligned}l^{-1}\hat{\Upsilon}^*_{bkk}(\hat{\beta}^*) &= \hat{\sigma}^*_{\epsilon\epsilon}(\hat{\beta}^*)\left\{\left(1-2\hat{\lambda}^*(\hat{\beta}^*)\right)\left(\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)P_Z\tilde{X}^*(\hat{\beta}^*)}{l} - \hat{\lambda}^*(\hat{\beta}^*)\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right)\right. \\ &\quad \left. +\hat{\lambda}^*(\hat{\beta}^*)\left(1-\hat{\lambda}^*(\hat{\beta}^*)\right)\frac{\tilde{X}^{*\prime}(\hat{\beta}^*)\tilde{X}^*(\hat{\beta}^*)}{l}\right\} \\ &= \sigma^b_{\epsilon\epsilon}\left\{(1-\lambda_n)\left[\bar{H}_{II,n} + \Sigma^b_{\tilde{V}\tilde{V}}\right]\right\} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\end{aligned}$$

Now we show the results for the non-normality adjustment terms $\hat{A}^*(\hat{\beta}^*)$ and $\hat{B}^*(\hat{\beta}^*)$. Let $\hat{a}^*(\hat{\beta}^*) = \left(\hat{\epsilon}_1^{*2}(\hat{\beta}^*) - \sigma_{\epsilon\epsilon}^b, ..., \hat{\epsilon}_n^{*2}(\hat{\beta}^*) - \sigma_{\epsilon\epsilon}^b\right)'$, $a^* = \left(\epsilon_1^{*2} - \sigma_{\epsilon\epsilon}^b, ..., \epsilon_n^{*2} - \sigma_{\epsilon\epsilon}^b\right)'$, and $\bar{V}^* = M_Z\tilde{V}^*$. By Z including a constant we have $n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{*2}(\hat{\beta}^*)\hat{V}_i^*(\hat{\beta}^*) = n^{-1}\hat{V}^{*\prime}(\hat{\beta}^*)\hat{a}^*(\hat{\beta}^*)$.

Note that $\hat{V}^*(\hat{\beta}^*) - \bar{V}^* = M_Z\left(Z\hat{\Pi} + \epsilon^*\frac{\sigma_{V\epsilon}^{b\prime}}{\sigma_{\epsilon\epsilon}^b} - \hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*\prime}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*\prime}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*)}\right) = M_Z\left(\epsilon^*\frac{\sigma_{V\epsilon}^{b\prime}}{\sigma_{\epsilon\epsilon}^b} - \hat{\epsilon}^*(\hat{\beta}^*)\frac{\hat{\epsilon}^{*\prime}(\hat{\beta}^*)X^*}{\hat{\epsilon}^{*\prime}(\hat{\beta}^*)\hat{\epsilon}^*(\hat{\beta}^*}\right.$ Then in Case (I), it follows by $n^{-1}\left\|\hat{\epsilon}^*(\hat{\beta}^*) - \epsilon^*\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$ and $\left\|\frac{\hat{\epsilon}^{*\prime}X^*}{\hat{\epsilon}^{*\prime}\hat{\epsilon}^*} - \frac{\sigma_{\epsilon V}^b}{\sigma_{\epsilon\epsilon}^b}\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$ that $n^{-1}\left\|\hat{V}^*(\hat{\beta}^*) - \bar{V}^*\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$. By using similar arguments, we obtain $n^{-1}\left\|\hat{a}^*(\hat{\beta}^*) - a^*\right\|^2 = O_{P^*}\left(\frac{1}{r_n}\right)$. Furthermore, notice that by using arguments similar to Lemma .3.1 and by Markov inequality $n^{-1}a^{*\prime}a^* = n^{-1}\sum_{i=1}^n\left(\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b\right)^2 = O_P(1)$; by using arguments similar to Lemma .3.2, we obtain $n^{-1}\bar{V}^{*\prime}\bar{V}^* = O_{P^*}(1)$. Thus, we obtain by Cauchy-Schwarz inequality that in Case (I)

$$\begin{aligned} n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{*2}(\hat{\beta}^*)\hat{V}_i^*(\hat{\beta}^*) - n^{-1}a^{*\prime}\bar{V}^* &= n^{-1}\left(\hat{a}^*(\hat{\beta}^*) - a^*\right)'\left(\hat{V}^*(\hat{\beta}^*) - \bar{V}^*\right) \\ &+ n^{-1}\left(\hat{a}^*(\hat{\beta}^*) - a^*\right)'\bar{V}^* + n^{-1}a^{*\prime}(\hat{V}^*(\hat{\beta}^*) - \bar{V}^*) = O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \end{aligned}$$

By Lemma .3.3, we have $n^{-1}a^{*\prime}\bar{V}^* = (1-\lambda_n)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right) + O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$. Then, it follows by the triangle inequality that

$$n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^{*2}(\hat{\beta}^*)\hat{V}_i^*(\hat{\beta}^*) = (1-\lambda_n)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \tag{15}$$

Now, let $d_i = \frac{1}{\sqrt{l}}(P_{ii} - \lambda_n)$ and $d = (d_1, ..., d_n)'$. Notice that $\|d\|^2 \leq 1$ and $E^*\left\|V^{*\prime}P_Zd\right\|^2 = O_P(1)d'd = O_P(1)$. Thus, $V^{*\prime}P_Zd = O_{P^*}(1)$ by Markov inequality. Then, $\frac{1}{\sqrt{r_n}}\sum_{i=1}^n\hat{\Gamma}_i^*\left(\frac{P_{ii}-\lambda_n}{\sqrt{l}}\right) = \frac{1}{\sqrt{r_n}}\left(\hat{\Pi}'Z'd + V^{*\prime}P_Zd\right) =$

$\frac{1}{\sqrt{r_n}}\hat{\Pi}'Z'd+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right)$, where $\hat{\Gamma}^*=P_ZX^*$. Then we obtain with eq.(15) that

$$r_n^{-1}\hat{A}^*(\hat{\beta}^*)=(1-\lambda_n)\left(\sqrt{\frac{l}{r_n}}\right)\sum_{i=1}^{n}\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{\sqrt{lr_n}}\right)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \tag{16}$$

For Case (II), we obtain $n^{-1}\left\|\hat{a}^*(\hat{\beta}^*)-a^*\right\|^2=O_{P^*}\left(\frac{1}{l}\right)$, $n^{-1}\left\|\hat{V}^*(\hat{\beta}^*)-\bar{V}^*\right\|^2=O_{P^*}\left(\frac{1}{l}\right)$, $n^{-1}a^{*\prime}\bar{V}^*=(1-\lambda_n)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right)+O_{P^*}\left(\frac{1}{\sqrt{n}}\right)$, $n^{-1}\sum_{i=1}^n\hat{\epsilon}_i^{*2}(\hat{\beta}^*)\hat{V}_i^*(\hat{\beta}^*)=(1-\lambda_n)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right)+O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, and

$$l^{-1}\hat{A}^*(\hat{\beta}^*)=(1-\lambda_n)\sum_{i=1}^{n}\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{l}\right)E^*\left(\epsilon_i^{*2}\tilde{V}_i^*\right)+O_{P^*}\left(\frac{1}{\sqrt{l}}\right)\to^{P^*}0,$$

in probability, because $\sum_{i=1}^n\left(\frac{(P_{ii}-\lambda_n)\hat{\Pi}'Z_i}{l}\right)\to^{P}0$.

For $\hat{B}^*(\hat{\beta}^*)$ term, it follows by using similar arguments as for $\hat{A}^*(\hat{\beta}^*)$ that for Case (I)

$$\left\|n^{-1}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*)-\hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*)-\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\right)\bar{V}_i^*\bar{V}_i^{*\prime}\right\|=O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right),$$

$$\left\|n^{-1}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*)-\hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*)\right)\left(\tilde{V}_i^*(\hat{\beta}^*)\tilde{V}_i^{*\prime}(\hat{\beta}^*)-\bar{V}_i^*\bar{V}_i^{*\prime}\right)\right\|=O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right).$$

by which we obtain

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*)-\hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*)\right)\tilde{V}^*(\hat{\beta}^*)\tilde{V}^{*\prime}(\hat{\beta}^*) &= n^{-1}\sum_{i=1}^{n}\left(\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b\right)\bar{V}_i^*\bar{V}_i^{*\prime}+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \\ &= (1-2\lambda_n+\lambda_n\phi_n)E^*\left((\epsilon_i^{*2}-\sigma_{\epsilon\epsilon}^b)\tilde{V}_i^*\tilde{V}_i^{*\prime}\right)+O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right) \end{aligned}$$

where the second equality follows from Lemma .3.3. It follows that

$$\hat{B}^*(\hat{\beta}^*) = (\phi_n - \lambda_n)E^*\left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b)\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(\frac{1}{\sqrt{r_n}}\right). \qquad (17)$$

Also, we have for Case (II)

$$n^{-1}\sum_{i=1}^{n}\left(\hat{\epsilon}_i^{*2}(\hat{\beta}^*) - \hat{\sigma}_{\epsilon\epsilon}^*(\hat{\beta}^*)\right)\tilde{V}^*(\hat{\beta}^*)\tilde{V}^{*'}(\hat{\beta}^*) = (1-2\lambda_n+\lambda_n\phi_n)E^*\left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b)\tilde{V}_i^*\tilde{V}_i^{*'}\right)+O_{P^*}\left(\frac{1}{\sqrt{l}}\right),$$

and $\hat{B}^*(\hat{\beta}^*) = (\phi_n - \lambda_n)E^*\left((\epsilon_i^{*2} - \sigma_{\epsilon\epsilon}^b)\tilde{V}_i^*\tilde{V}_i^{*'}\right) + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$.

Finally, we show the result for t_{cse}^*, the bootstrap analogue of the t-ratio based on the CSE. Notice that in Case (I)

$$\begin{aligned} r_n\hat{\Lambda}^*(\hat{\beta}^*) &= \left(r_n^{-1}\hat{H}^*(\hat{\beta}^*)\right)^{-1}\left\{r_n^{-1}\left(\hat{\Upsilon}_{bkk}^*(\hat{\beta}^*) + \hat{A}^*(\hat{\beta}^*) + \hat{A}^{*'}(\hat{\beta}^*) + \hat{B}^*(\hat{\beta}^*)\right)\right\}\left(r_n^{-1}\hat{H}^*(\hat{\beta}^*)\right)^{-1} \\ &\to^{P^*} \bar{\Lambda}_I \end{aligned}$$

in probability, which follows from Lemma .3.6, Lemma .3.7, previous results for $\hat{\Upsilon}_{bkk}^*(\hat{\beta}^*)$, $\hat{A}^*(\hat{\beta}^*)$ and $\hat{B}^*(\hat{\beta}^*)$, the fact that $\hat{\beta}^* - \hat{\beta} = O_{P^*}\left(1/\sqrt{r_n}\right)$ in Case (I), and by using arguments similar to those in Theorem 3.3.1. It follows that

$$t_{cse}^* = \frac{c'\sqrt{r_n}\left(\hat{\beta}^* - \hat{\beta}\right)}{\sqrt{c'r_n\hat{\Lambda}^*\left(\hat{\beta}^*\right)c}} \to^{d^*} N(0,1)$$

in probability, by Theorem 3.3.1 and by continuous mapping theorem for weak convergence in probability (e.g., Xiong and Li (2008), Theorem 3.1).

For Case (II), we have $l\hat{\Lambda}^*(\hat{\beta}^*) = \bar{\Lambda}_{II} + O_{P^*}\left(\frac{1}{\sqrt{l}}\right)$, and the result follows by using similar arguments as in Case (I). ■