

Automatic Inference for Value-Added Regressions*

Tian Xie

Department of Economics, University College London

November 29, 2025

[Click here for the latest version](#)

Abstract

A large empirical literature regresses outcomes on empirical Bayes shrinkage estimates of value-added, yet little is known about whether this approach leads to unbiased estimates and valid inference for the downstream regression coefficients. We study a general class of empirical Bayes estimators and the properties of the resulting regression coefficients. We show that estimators can be asymptotically biased and inference can be invalid if the shrinkage estimator does not account for heteroskedasticity in the noise when estimating value added. By contrast, shrinkage estimators properly constructed to model this heteroskedasticity perform an automatic bias correction: the associated regression estimator is asymptotically unbiased, asymptotically normal, and efficient in the sense that it is asymptotically equivalent to regressing on the true (latent) value-added. Further, OLS standard errors from regressing on shrinkage estimates are consistent in this case. As such, efficient inference is easy for practitioners to implement: simply regress outcomes on shrinkage estimates of value-added that account for noise heteroskedasticity.

Key words: empirical Bayes, shrinkage estimators, teacher value-added, errors in variables.

JEL classification codes: C12.

*I thank Tim Christensen, Raffaella Giacomini, and Liyang Sun for continuous guidance and generous support throughout this project. I would also like to thank Luis Alvarez, Debopam Bhattacharya, Kirill Borusyak, Pedro Carneiro, Jiafeng Chen, Xiaohong Chen, Andrew Chesher, Max Cytrynbaum, Ben Deaner, Aureo de Paula, Hugo Freeman, Christophe Gaillac, Sukjin Han, Stephen Hansen, Toru Kitagawa, Yuichi Kitamura, Roger Koenker, Dennis Kristensen, Soonwoo Kwon, Simon Lee, Attila Lindner, Oliver Linton, Eric Mbakop, Lars Nesheim, Alexei Onatski, Matthias Parey, Kirill Ponomarev, Silvia Sarpietro, Michela Tincani, Gabriel Ulyssea, Edward Vytlačil, Guanyi Wang, Weining Wang, Ming Yang, Andrei Zeleneev, and participants of the IAAE Conference (2025), Bristol Econometric Study Group Conference (2025), 2nd UCL-CeMMAP Ph.D. Econometrics Research Day, Midwest Econometrics Group Conference (2025), 4th International Econometrics PhD Conference Erasmus University Rotterdam, and seminars at UCL, Yale University, and University of Cambridge. All errors are my own. Email address: tian.xie.20@ucl.ac.uk.

1 Introduction

Empirical Bayes shrinkage estimators are widely applied in settings with a large number of latent individual effects observed via noisy measurements. Economic applications include the study of teachers' value-added to student test scores (e.g. [Chetty, Friedman, and Rockoff, 2014a,b](#)), location effects (e.g. [Chetty and Hendren, 2018](#)), hospital quality (e.g. [Hull, 2018](#)), nursing home quality (e.g. [Einav, Finkelstein, and Mahoney, 2025](#)), firm-level discrimination (e.g. [Kline, Rose, and Walters, 2022](#)), among others. Shrinkage estimates are typically not the final object, but instead serve as inputs to downstream analyses. A typical downstream use is to regress other variables of economic interest on the estimated individual effects.

It is common for researchers to treat the shrinkage estimates as if they were the true individual effects when performing conventional inference in the downstream regression. This widespread empirical approach implicitly assumes the property of automatic inference, i.e., no adjustment is required to achieve nominal coverage rates of confidence intervals. However, empirical Bayes shrinkage estimates generally differ from the latent true effects, introducing measurement error. Moreover, there is also the potential for a generated regressor problem ([Pagan, 1984](#)), because the regressors are themselves generated in a first-stage estimation step. Thus, there may be threats to the validity of both the estimator (via measurement error bias), and inference (via incorrect standard errors).

In empirical applications, various linear shrinkage estimators have been used as regressors. A leading example is the literature studying how teachers' value-added for test scores impacts students' future outcomes. Data on students' test scores are aggregated into shrinkage estimates, which are constructed as a weighted average of each teacher's individual mean score and all teachers' pooled mean score. Most studies employ individualized shrinkage, where noisier estimates (e.g., from smaller classes) are shrunk more aggressively toward the pooled mean. Empirical work using this approach includes [Jacob and Lefgren \(2007\)](#), [Jacob and Lefgren \(2008\)](#), [Kane and Staiger \(2008\)](#), [Chandra, Finkelstein, Sacarny, and Syverson \(2016\)](#), [Jackson \(2018\)](#), [Abdulkadiroğlu, Pathak, Schellenberg, and Walters \(2020\)](#), [Bau and Das \(2020\)](#), [Biasi and Sarsons \(2022\)](#), [Warnick, Light, and Yim \(2024\)](#), [Andrabi, Bau, Das, and Khwaja \(2025\)](#), and [Angelova, Dobbie, and Yang \(2025\)](#). Individualized shrinkage differs from equal shrinkage approaches, such as [Chetty et al. \(2014a,b\)](#), which apply a uniform

shrinkage rule to all units. However, the theoretical impact of these different shrinkage schemes on downstream estimates and inference has not been studied to date.

This paper is the first to formally study the impact of different shrinkage schemes on the properties of downstream regressions, and to characterize when automatic inference does or does not obtain. We focus on linear shrinkage estimators, which are widely used in practice due to their simplicity and ease of implementation. We develop a general econometric framework for analyzing a broad class of individualized shrinkage estimators as regressors. These estimators are weighted averages whose weights reflect signal-to-noise ratios estimated from the data. They can be classified as individual-weight (individualized shrinkage) or common-weight (equal shrinkage). Our central finding shows that automatic inference hinges on the treatment of heteroskedastic measurement error. In particular, failing to account for heteroskedasticity in the shrinkage weights leads to *invalid* downstream inference, with asymptotically biased estimators and coverage rates from conventional confidence intervals below nominal coverage. By contrast, automatic inference obtains if the weights properly account for this heteroskedasticity. In this case, the conventional confidence intervals achieve nominal coverage without further correction, and the resulting downstream coefficient estimator is asymptotically equivalent to the infeasible OLS regression on the true latent effects. Thus, there is no efficiency loss from regressing on the estimated individual effects instead of the true individual effects.

These findings have important practical implications for current practice. Many empirical implementations model heteroskedastic measurement error solely by the number of measurements per individual, e.g., class sizes. Yet idiosyncratic noise variances may also differ across individuals. We show that accounting only for heterogeneity in the number of measurements can yield invalid inference if noise variances are correlated with the number of measurements. This is because the individualized adjustments systematically under- or over-correct depending on the degree of heterogeneity. This, in turn, introduces a non-classical measurement error, where the bias can lead to amplification rather than attenuation. Inference remains valid when the two sources of heterogeneity are independent. Properly accounting for heteroskedasticity in the shrinkage weights ensures robustness to such dependence and allows automatic inference for the downstream regression coefficient.

Accordingly, our analysis provides straightforward practical guidance for implementation. In the teacher value-added example, the weights may vary across teachers

and are determined by the estimated signal-to-noise ratio. Components of the ratio are estimated by within- and across-teacher variances, which properly account for heteroskedasticity in the measurement error. In predicting long-run outcomes as a function of latent value-added, we simply regress the outcomes on these shrinkage estimates. From the reported results, conventional OLS standard errors and confidence intervals are valid for inference.

Our results do not impose distributional assumptions and have a clear intuition. Linear shrinkage estimators are the empirical Bayes posterior mean under normality of both the measurement error and the latent individual effect (Efron and Morris, 1973; Morris, 1983). However, while the functional form of shrinkage is derived under normality, we study regressions in a semiparametric context and none of our results rely on normality. The overall intuition is as follows. In the presence of noise, it is well known that regressions on unshrunk individual means suffer from a classical measurement error problem. Shrinkage estimators reduce the variability of the estimated individual effects and thereby offset the variance inflation that comes from noise. What is not clear is whether they offset it by just the right amount, so that downstream regression estimators are asymptotically unbiased and downstream inference is valid. We show that for some scenarios this is the case, while for others it is not.

To formalize our analysis, we first restate the established baseline result that the unshrunk individual mean suffers from classical errors-in-variables, resulting in attenuation bias and invalid inference when used as a downstream regressor. We then show our main results on individual-weight shrinkage. First, we outline the precise conditions under which inference fails, even when heterogeneity in the number of measurements is accounted for. Second, we establish how a properly constructed, fully individualized shrinkage estimator achieves asymptotic unbiasedness, inferential validity, and asymptotic efficiency in the downstream regression. Finally, we analyze the common-weight estimator as a point of comparison. Because of its close connection to conventional bias-correction methods, this estimator also addresses the inference problem. Nevertheless, we focus our attention on the individual-weight approach because it is more used in contemporary work and no theoretical results currently exist to justify its use in downstream regressions. While our focus has been on individual-weight shrinkage that assumes (as is standard) independence between individual effects and variance of measurement error, our results can be generalized to

relax this assumption. Recent work has derived empirical Bayes estimators allowing such dependence (Chen, 2023), but has not used them in downstream regressions. In on-going work, we extend our analysis to this new class of estimators and show that they also yield automatic inference for downstream regressions.

We conduct Monte Carlo simulations to demonstrate the finite-sample performance of the different shrinkage estimators when used as regressors. The results confirm our asymptotic theory. The unshrunk individual means yield substantial attenuation bias and under-coverage of confidence intervals. Individual-weight shrinkage that accounts only for heterogeneity in the number of measurements can also lead to bias and under-coverage when noise variances are correlated with the number of measurements. In contrast, heteroskedastic individual-weight shrinkage that accounts for both sources of heterogeneity yields valid coverage, and achieves very close performance to regressing on the true latent individual effects. Common-weight shrinkage also performs well, but is outperformed by heteroskedastic individual-weight shrinkage in finite samples when the sample size is moderate.

We illustrate our results using two empirical applications. The first revisits the firm discrimination data of Kline et al. (2022), while the second is based on the school value-added data of Andrabi et al. (2025). In the first application, we consider a regression of future discrimination levels on estimated discrimination at the firm level. We find that individual-weight heteroskedastic shrinkage yields confidence intervals more strongly supportive of a positive predictive effect. In the second application, we regress private school fees on school value-added estimates. We find that individual-weight heteroskedastic shrinkage produces slightly different confidence intervals from those reported in the original study, and it reaffirms the positive effect of school value-added on private school fees.

Related Literature This paper relates to an older literature that studies the use of common-weight shrinkage estimators as downstream regressors. Regressing on common-weight shrinkage estimators can be traced back to Whitemore (1989), who demonstrates via simulations that common-weight shrinkage yields consistent downstream coefficients. That paper also points out informally its equivalence to bias-correction methods. Building on this insight, Guo and Ghosh (2012) provide theoretical justifications regarding the quadratic risk reduction of such estimators for downstream coefficients. Chetty et al. (2014a,b) construct common-weight shrink-

age estimators based on the best linear predictor, emphasizing that the estimator is bias-free when used as the regressor. Their estimator can be viewed as equivalent to IV methods, and we unify this IV perspective along with the bias-correction perspective of common-weight shrinkage in Section 3.4. Deeb (2021) further exploits the equivalence to IV methods and develops inference results for Chetty et al. (2014a,b)’s estimator, highlighting the need to adjust conventional OLS standard errors to account for errors in estimation of individual effects and nuisance parameters. While this line of work clarifies inference in the common-weight shrinkage setting, it essentially relies on the equivalence to IV and does not extend to individual-weight shrinkage. These results are not applicable to the analysis of individual-weight shrinkage, which we study.

We formalize and extend prior work on individual-weight shrinkage. Consistency of regression estimators based on individual-weight shrinkage has been briefly discussed in Jacob and Lefgren (2007), Kane and Staiger (2008), Abdulkadiroğlu et al. (2020), Walters (2024), and Andrabi et al. (2025). In particular, Walters (2024) highlights independence between individual effects and the variance of the noise as a necessary condition for using widespread individualized shrinkage strategies. Beyond such consistency results, there has been no work studying the consequences for downstream inference of using individual-weight shrinkage estimators as regressors. We show that valid inference depends critically on how heteroskedasticity is treated in the shrinkage weights. We establish asymptotic unbiasedness, asymptotic normality, and consistency of standard errors for the downstream regression estimator based on properly constructed individual-weight shrinkage.

This paper is also related to non-shrinkage options for various downstream models. Chang, Huang, Chen, and Liao (2024) leverage information from the estimated prior to develop correction methods for nonlinear downstream models. For regressions, this literature strand is largely concerned with bias-correction. As noted from the equivalence, discussions from this category also apply to common-weight shrinkage. Kline, Saggio, and Sølvssten (2020) and Bonhomme and Denis (2024) emphasize correct estimation of moments for value-added that solves the errors-in-variables problem. Chen, Gu, and Kwon (2025) show that standard bias-correction methods for the downstream regression remain consistent even if the individual effects and measurement error are correlated. We contribute to this literature by establishing conditions under which conventional shrinkage estimators yield valid inference for downstream

regressions when individual effects and noise are independent. In ongoing work, we extend the analysis to this correlated case.

Outline This paper proceeds as follows. Section 2 introduces the framework and its implications. Section 3 discusses a broad class of shrinkage estimators, establishes their asymptotic normality and the consistency of standard errors, and discusses the validity of standard inference approaches for the downstream regression. It also provides some practical implementation guidelines. Section 4 reports Monte Carlo simulations illustrating the finite-sample properties of the estimators. These results reinforce our main findings that properly constructed individualized shrinkage estimators yield valid inference, while improperly constructed ones do not. Section 5 revisits the study of Kline et al. (2022) on labor market discrimination. Section 6 revisits the study of Andrabi et al. (2025) on school value-added. Finally, Section 7 concludes. An appendix contains the additional lemmas, proofs, and extensions.

2 Setup

Each unit i is associated with a latent individual effect θ_i , which we will refer to in what follows as value-added. Let $\boldsymbol{\theta} := (\theta_i)_{i=1}^n$ denote the vector of these latent effects for the n sampled units. A large body of research studies the relationship between value-added θ_i and some outcome of economic interest Y_i . To fix ideas, consider the example of Kane and Staiger (2008). In that paper, θ_i is teacher i 's latent value-added for students' test scores, and Y_i is the average long-term test score outcome of students taught by teacher i . The relationship is studied using the regression model

$$Y_i = \alpha + \beta\theta_i + u_i, \tag{1}$$

where the coefficient β captures the downstream effect. It is assumed that the error term u_i has mean zero and is uncorrelated with θ_i , $\mathbb{E}[u_i\theta_i] = 0$. Research questions about how teacher value-added relates to long term outcomes such as college attendance or earnings can be addressed by performing inference on β .

Inference on β faces the challenge that $\boldsymbol{\theta}$ is unobserved, so the parameter of interest β cannot be estimated by regressing Y_i on θ_i . Instead, for each unit i we only have data $(Y_i, X_i) := (Y_i, X_{i,1}, \dots, X_{i,J_i})$, $i = 1, \dots, n$, where Y_i is the outcome,

$X_{i,1}, \dots, X_{i,J_i}$ are repeated measurements of θ_i , and J_i is the number of observations available for estimating θ_i . Specifically, $X_i \in \mathbb{R}^{J_i}$ are noisy repeated measurements for θ_i from the model

$$X_{i,j} = \theta_i + \epsilon_{i,j}, \quad (2)$$

where $\epsilon_{i,j}$ is the noise term. In the example, $X_{i,j}$ is the test score of student j taught by teacher i . The individual mean score for teacher i is denoted by \bar{X}_i , representing the average test score within teacher i 's students. The pooled mean score is denoted by \bar{X} , representing the grand average across all teachers and students. In this paper, we primarily focus on the case where the noise is independent of the value-added, i.e., $\epsilon_{i,j} \perp \theta_i$.

3 Shrinkage Estimators as Regressors

In this section, we analyze the inferential properties of different shrinkage estimators when used as downstream regressors. We follow the standard two-step workflow in empirical practice: Step 1: Estimate value-added θ_i with a shrinkage estimator $\hat{\theta}_i$; Step 2: Regress Y_i on $\hat{\theta}_i$ and report conventional OLS estimates, standard errors and confidence intervals (i.e., treating the $\hat{\theta}_i$ as regular data rather than first-stage estimates). Our interest is in the validity of inference in Step 2 when different shrinkage estimators are used in Step 1.

The primary implication of our asymptotic framework is that the measurement error in \bar{X}_i does not vanish relative to the sampling error for β , mimicking the finite-sample problem faced in practice, where both play a role. This persistence of measurement error is the central econometric challenge we address. As we will show in Section 3.1, this asymptotic setup confirms that the naive unshrunk estimator (simply regressing Y_i on \bar{X}_i) suffers from classical errors-in-variables bias and yields invalid inference. This provides a useful baseline and motivation for the other estimators.

We consider four primary classes of estimators for $\hat{\theta}$, which we analyze in the order of our main results. For each class, we derive the asymptotic distribution of the OLS estimator of β in the downstream regression. In Section 3.1, we begin with the unshrunk fixed-effect (FE) estimator \bar{X}_i as a baseline. We then move on to individual-weight shrinkage. We distinguish between two forms: homoskedas-

tic individual-weight shrinkage (HO), which models heterogeneity in measurement precision solely by the number of measurements (J_i), and heteroskedastic individual-weight shrinkage (HE), which additionally models idiosyncratic noise variances. We study homoskedastic individual-weight shrinkage in Section 3.2 and heteroskedastic individual-weight shrinkage in Section 3.3. Finally, to connect with other approaches in the literature, we analyze common-weight shrinkage (CW), which applies a single, uniform weight across units.

3.1 Fixed Effects (FE)

As a baseline, we first discuss using the fixed-effect estimator (FE) with no shrinkage. In this estimator, value-added for unit i is simply estimated by the individual average: $\hat{\theta}_{i,\text{FE}} = \bar{X}_i$, for $i = 1, \dots, n$.

As we show formally below, using \bar{X}_i as a regressor suffers from the classical errors-in-variables (EIV) problem, leading to attenuation bias. This bias shifts the location of standard OLS confidence intervals for β away from the truth, rendering them invalid for inference. In the following, we first state and discuss assumptions that are needed for the asymptotic properties.

Assumption 3.1. 1. J_i is independent of θ_i , u_i and $\epsilon_{i,j}$, and $J_i \geq 3$, a.s.

2. $\mathbb{E}[u_i] = 0$, $\mathbb{E}(u_i\theta_i) = 0$. $\mathbb{E}[Y_i^4] < \infty$.

3. $\mathbb{E}[\epsilon_{i,j} | \sigma_i^2] = 0$, $\mathbb{E}[\epsilon_{i,j}^2 | \sigma_i^2] = \sigma_i^2$, $\mathbb{E}[|\epsilon_{i,j}|^L | \sigma_i^2] \leq K\sigma_i^L$, for some $L \geq 3$. Also, $\epsilon_{i,j} \perp\!\!\!\perp \theta_i$.

4. $u_i \perp\!\!\!\perp \epsilon_{i,j} | \theta_i$.

5. $\mathbb{E}[\theta_i^4] < \infty$.

6. $\mathbb{E}[\sigma_i^{16}] < \infty$.

7. $\sqrt{n}\mathbb{E}[J_i^{-1}] \rightarrow \kappa$, $\kappa \in [0, +\infty)$.

8. $\sqrt{n}\mathbb{E}[J_i^{-2}] \rightarrow 0$.

Assumption 3.1.1 assumes independent numbers of measurements J_i . In Appendix D, the condition will be relaxed to allow dependence between J_i and σ_i^2 ,

accompanied by slightly stronger assumptions than Assumption 3.1.7 and Assumption 3.1.8. We keep the independence for ease of exposition. Assumption 3.1.2 imposes unconditional exogeneity of the true θ_i and u_i , and moment conditions on Y_i .

Assumption 3.1.3 places moment conditions on the noise $\epsilon_{i,j}$. The moment conditions are general and cover a wide range of distributions for $\epsilon_{i,j}/\sigma_i$, including normal, bounded, and even some asymmetric distributions. Thus, while normality of \bar{X}_i is assumed to derive the functional form of parametric empirical Bayes estimators, the inference results we derive do not require any normality assumption. Independence $\epsilon_{i,j} \perp \theta_i$ is standard in the empirical Bayes literature to justify shrinkage estimators without covariates. Assumption 3.1.4 excludes further effects of measurement errors on Y_i conditional on θ_i . Assumption 3.1.5 and Assumption 3.1.6 are standard moment conditions for technical arguments.

Finally, Assumption 3.1.7 and Assumption 3.1.8 specify the asymptotic framework. Note that if we have the same number of measurements, $J_i = J$, then Assumption 3.1.8 is implied by Assumption 3.1.7. Thus Assumption 3.1.7 is the essential condition. Proposition 3.1 below holds without Assumption 3.1.8. The asymptotic framework has three motivations. First, it aligns with the standard justification for normality of \bar{X}_i via the central limit theorem with J_i reasonably large (Walters, 2024). Second, it reflects the data structure in common applications, such as the teacher value-added setting, where J_i often has a similar magnitude to \sqrt{n} . In these settings, the number of teachers is in the thousands and the number of students per teacher is in the tens. For instance, in the study of North Carolina data in Deeb (2021), the total number of teachers $n = 5266$, and the total number of students is 388,191, so we would expect J_i to be on average about 74, which is comparable to $\sqrt{n} \approx 73$. In another example from Bau and Das (2020), $\sqrt{n} \approx 39$ and J_i is on average about 15. Third, it provides useful approximations to the finite-sample problem faced by researchers. The product $\sqrt{n}\mathbb{E}[J_i^{-1}]$ quantifies the relative scale of measurement error in the estimates of θ_i to sampling error in β . When $\kappa = 0$, J_i grows asymptotically much larger than \sqrt{n} , implying that measurement error in θ_i is negligible for the purpose of inference in the regression. When $\kappa > 0$, measurement error in θ_i is of a similar order to the sampling error. In the previous two application examples, we have $\hat{\kappa}_1 \approx 73/74 \approx 1$, and $\hat{\kappa}_2 \approx 39/15 \approx 2.6$,¹ as sample counterparts

¹These approximations are lower bounds, because Jensen’s inequality produces a larger value if J_i varies across i

of κ . A finite positive κ is thus appropriate for the context in which tens of students per teacher and thousands of teachers coexist.

Our asymptotic framework is related to the small-variance approximation in Chesher (1991) and Evdokimov and Zeleneev (2019, 2023). This literature does not put a structure on the form of the measurement error, but assumes its variance is proportional to $n^{-1/2}$. Here we know that the measurement error from estimation of θ_i has a variance proportional to J_i^{-1} , which leads to Assumption 3.1.7. It is also standard in settings with both cross-sectional and individual-specific datasets, such as large N, T panels (Pesaran, 2006), factor augmented regressions (Gonçalves and Perron, 2014), and unstructured data (Battaglia, Christensen, Hansen, and Sacher, 2024).

Denote by $\hat{\beta}_{\text{FE}}$ the OLS estimator from the regression of Y_i on $\hat{\theta}_{i,\text{FE}}$ using observations $i = 1, \dots, n$. The following result implies that conventional OLS inference on β using $\hat{\beta}_{\text{FE}}$ is invalid.

Proposition 3.1. *Suppose Assumption 3.1 holds. Then we have*

$$\sqrt{n} \left(\hat{\beta}_{\text{FE}} - \beta \right) \rightarrow_d N \left(-\kappa \beta \frac{\mathbb{E}[\sigma_i^2]}{\text{Var}(\theta_i)}, \frac{\mathbb{E}[u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2]}{(\text{Var}(\theta_i))^2} \right).$$

In the special case where $\kappa = 0$ (meaning measurement error is asymptotically negligible), the bias term disappears and the asymptotic distribution of $\sqrt{n} \left(\hat{\beta}_{\text{FE}} - \beta \right)$ is centered at zero. In our primary framework with $\kappa > 0$, $\hat{\beta}_{\text{FE}}$ is consistent and asymptotically normal, with the efficient asymptotic variance. Thus, there is no generated regressor problem: OLS standard errors are consistent, following similarly to Theorem 3.4 in Section 3.3. However, its asymptotic distribution is centered away from zero due to attenuation bias. This has important practical consequences. Consider the conventional confidence interval for β , formed as $\hat{\beta}_{\text{FE}} \pm 1.96 \times \text{SE}(\hat{\beta}_{\text{FE}})$ where $\text{SE}(\hat{\beta}_{\text{FE}})$ represents the usual OLS standard error. This confidence interval will be centered away from β , and consequently will not have valid coverage for β . To give a sense of the magnitude of under-coverage, the simulations in Section 4 show that the coverage of a 95% confidence interval is only about 70%.

3.2 Homoskedastic Individual-Weight Shrinkage (HO)

In this subsection, we discuss the individual-weight shrinkage when the variance of \bar{X}_i is assumed to only depend on the number of measurements J_i . That is, $\text{Var}(\bar{X}_i | \theta_i) = \sigma^2/J_i$. Estimators constructed under this assumption are referred to in this paper as homoskedastic individual-weight shrinkage (HO), since the noise $\epsilon_{i,j}$ is homoskedastic across i (note that we still allow for heteroskedasticity in the downstream regression, however). Here the primary source of heteroskedasticity is believed to arise from variation in the number of measurements across units i . The functional form of the shrinkage weights takes this heterogeneity into account.

Following the estimators proposed by [Jacob and Lefgren \(2007\)](#) and [Kane and Staiger \(2008\)](#), the homoskedastic individual-weight shrinkage estimator θ is constructed as

$$\hat{\theta}_{i,\text{HO}} := \frac{\hat{\sigma}_\theta^2}{\frac{1}{J_i}\hat{\sigma}^2 + \hat{\sigma}_\theta^2} \bar{X}_i + \frac{\frac{1}{J_i}\hat{\sigma}^2}{\frac{1}{J_i}\hat{\sigma}^2 + \hat{\sigma}_\theta^2} \bar{X}.$$

The estimator is the empirical Bayes posterior mean under normality of θ_i and $\epsilon_{i,j}$. The weight for \bar{X}_i is $w_{i,\text{HO}}$, obtained by replacing the oracle shrinkage weight $\frac{\text{Var}(\theta_i)}{\sigma^2/J_i + \text{Var}(\theta_i)}$ with its empirical counterparts. These variances are estimated as

$$\begin{aligned} \hat{\sigma}^2 &:= \widehat{\text{Var}}(X_{i,j} - \bar{X}_i), \\ \hat{\sigma}_\theta^2 &:= \widehat{\text{Cov}}(\bar{X}_{i,t}, \bar{X}_{i,t-1}), \end{aligned}$$

where $\bar{X}_{i,t}, \bar{X}_{i,t-1}$ are subset averages from splitting the measurements for each i into two subsets. In a context with time periods, they can also be averages for two time periods. Note that the same $\hat{\sigma}^2$ is applied to units, with variation in J_i reflecting differences in measurement precision.

We now derive the asymptotic distribution of the OLS estimator $\hat{\beta}_{\text{HO}}$ of regressing Y_i on $\hat{\theta}_{i,\text{HO}}$ in two cases. In the first case, we assume the homoskedasticity in noise holds (i.e., $\sigma_i^2 = \sigma^2$ for all i). In the second case, we allow for heteroskedasticity in noise (i.e., σ_i^2 varies across i).

If homoskedasticity holds, this method is expected to correctly model the signal-to-noise ratio. Indeed, as we show below in [Proposition 3.2](#), regressing Y_i on $\hat{\theta}_{i,\text{HO}}$ yields asymptotic unbiasedness and valid inference on β under homoskedasticity.

If the true DGP is heteroskedastic, we demonstrate problems in inference when

homoskedastic weights are applied. Here, denote the variance of $\bar{X}_i \mid \theta_i$ by σ_i^2/J_i . Regressing Y_i on $\hat{\theta}_{i,\text{HO}}$ can lead to invalid inference on β under noise heteroskedasticity. A leading case is when the number of measurements J_i is correlated with the variance σ_i^2 . In that case, it is natural to have J_i endogenously increased for observations with large σ_i^2 . But in this case the downstream regression estimator has an asymptotic bias which invalidates the standard approach to inference. The following result substantiates this claim, with details and proofs in Appendix D.

Proposition 3.2. *Suppose Assumption D.1 and D.2 hold, which allow for correlation between J_i and σ_i^2 . Then there exist DGPs with $\gamma > 0$ in which*

$$\sqrt{n} \left(\hat{\beta}_{\text{HO}} - \beta \right) \rightarrow_d N \left(\gamma\beta, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{(\text{Var} (\theta_i))^2} \right).$$

If instead J_i and σ_i^2 are independent, then

$$\sqrt{n} \left(\hat{\beta}_{\text{HO}} - \beta \right) \rightarrow_d N \left(0, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{(\text{Var} (\theta_i))^2} \right).$$

Proposition 3.2 clarifies the conditions required for regressing Y_i on $\hat{\theta}_{i,\text{HO}}$ to yield valid inference. While inference is valid if the homoskedasticity assumption holds or if J_i and σ_i^2 are independent, the estimator's performance is sensitive to violations of homoskedasticity. As $\gamma > 0$ in the bias term, the asymptotic bias is *amplification* instead of classical EIV attenuation bias. Generally, the sign and magnitude of γ depend on the dependence of J_i and σ_i^2 , and the bias can thus be positive or negative. It motivates the heteroskedastic estimator we consider next, which is designed to be robust to such dependence.

3.3 Heteroskedastic Individual-Weight Shrinkage (HE)

In this subsection, we discuss the heteroskedastic individual-weight (HE) estimator under noise heteroskedasticity. This estimator is designed to accommodate heterogeneity from both the number of measurements J_i and the idiosyncratic noise variance σ_i^2 . It improves upon $\hat{\theta}_{i,\text{HO}}$.

The heteroskedastic individual-weight shrinkage estimator is defined as

$$\begin{aligned}\hat{\theta}_{i,\text{HE}} &:= w_{i,\text{HE}}\bar{X}_i + (1 - w_{i,\text{HE}})\bar{X}, \text{ where} \\ w_{i,\text{HE}} &:= \frac{\hat{V}}{\frac{1}{J_i}\hat{\sigma}_i^2 + \hat{V}}.\end{aligned}\tag{3}$$

Here, \hat{V} is the estimator of $\text{Var}(\theta_i)$, and $\hat{\sigma}_i^2$ is the estimator of $\text{Var}(X_{i,j} | \theta_i)$. The variance estimators are constructed following [Kline et al. \(2020\)](#) and [Kline et al. \(2022\)](#):

$$\begin{aligned}\hat{\sigma}_i^2 &:= \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - \bar{X}_i)^2, \\ \hat{V} &:= \frac{1}{n} \sum_{k=1}^n (\bar{X}_k - \bar{X})^2 - \frac{n-1}{n^2} \sum_{k=1}^n \frac{1}{J_k} \hat{\sigma}_k^2.\end{aligned}$$

Unlike $\hat{\theta}_{i,\text{HO}}$, the weights $w_{i,\text{HE}}$ allow $\text{Var}(X_{i,j} | \theta_i)$ to differ across units, thereby ensuring robustness to heteroskedasticity.

Let $\hat{\beta}_{\text{HE}}$ denote the downstream OLS estimator from regressing Y_i on $\hat{\theta}_{i,\text{HE}}$. We take an intermediate step in deriving the properties of $\hat{\beta}_{\text{HE}}$. The estimator $\hat{\theta}_{i,\text{HE}}$ is an empirical Bayes estimator, with the estimated prior variance \hat{V} and prior mean \bar{X} for a normal prior on θ_i . We start by studying the property of OLS estimator $\hat{\beta}_{c,\text{HE}}$ from regressing Y_i on the semi-oracle estimator $\hat{\theta}_{i,c,\text{HE}}$ (with prior mean still estimated from the data),

$$\begin{aligned}\hat{\theta}_{i,c,\text{HE}} &:= c_i\bar{X}_i + (1 - c_i)\bar{X}, \text{ where} \\ c_i &:= \frac{\text{Var}(\theta_i)}{\frac{1}{J_i}\hat{\sigma}_i^2 + \text{Var}(\theta_i)}.\end{aligned}$$

Lemma [B.3](#) shows that the estimated variance \hat{V} is a consistent estimator for the true variance $\text{Var}(\theta_i)$. Thus, we build the properties of $\hat{\beta}_{\text{HE}}$ based on those of $\hat{\beta}_{c,\text{HE}}$. The following result shows that the OLS estimator $\hat{\beta}_{c,\text{HE}}$ from regression Y_i on $\hat{\theta}_{i,c,\text{HE}}$ is asymptotically normal, correctly centered, with the efficient asymptotic variance from the infeasible regression of Y_i on θ_i .

Lemma 3.1. *Suppose Assumption 3.1 holds. Then we have*

$$\sqrt{n} \left(\hat{\beta}_{c,\text{HE}} - \beta \right) \rightarrow_d N \left(0, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{(\text{Var} (\theta_i))^2} \right).$$

Lemma 3.1 shows that regressing Y_i on $\hat{\theta}_{i,c,\text{HE}}$ leads to an asymptotically unbiased estimator of β in infeasible cases where $\text{Var} (\theta_i)$ is known. The estimator performs asymptotically equivalently to that from regression of Y_i on θ_i . Next, we move on to the feasible estimator $\hat{\theta}_{i,\text{HE}}$.

The following result builds on Lemma 3.1 and establishes the asymptotic distribution if we use the feasible shrinkage estimator $\hat{\theta}_{i,\text{HE}}$. Just like $\hat{\beta}_{c,\text{HE}}$, the OLS estimator $\hat{\beta}_{\text{HE}}$ from regressing Y_i on $\hat{\theta}_{i,\text{HE}}$ is also asymptotically normal, correctly centered, with the efficient asymptotic variance.

Theorem 3.3. *Suppose Assumption 3.1 holds. Then we have that $\hat{\beta}_{\text{HE}}$ and $\hat{\beta}_{c,\text{HE}}$ are first-order asymptotically equivalent:*

$$\sqrt{n} \left(\hat{\beta}_{\text{HE}} - \beta \right) = \sqrt{n} \left(\hat{\beta}_{c,\text{HE}} - \beta \right) + o_p(1),$$

and so

$$\sqrt{n} \left(\hat{\beta}_{\text{HE}} - \beta \right) \rightarrow_d N \left(0, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{(\text{Var} (\theta_i))^2} \right).$$

Theorem 3.3 establishes that the OLS estimator based on the feasible regressor $\hat{\theta}_{i,\text{HE}}$ is first-order asymptotically equivalent to the estimator based on the semi-oracle $\hat{\theta}_{i,c,\text{HE}}$. Moreover, there is no efficiency loss relative to the infeasible regression of Y_i on the latent θ_i . The asymptotic variance is the efficiency bound for estimating β in models where θ_i is observed (See the unconditional moment case of Chamberlain, 1987). The asymptotic distribution doesn't depend on the value of κ . Thus, the OLS regression estimator $\hat{\beta}_{\text{HE}}$ is robust to the amount of measurement error (under the conditions of Theorem 3.3).

Remark 3.1. Recall from Proposition 3.1 that the asymptotic distribution of the estimator $\hat{\beta}_{\text{FE}}$ from regressing on unshrunk $\hat{\theta}_{i,\text{FE}}$ was biased by the κ term, leading to invalid inference. Here, even though we remain in the same $\kappa > 0$ regime where measurement error is persistent, that bias term is absent. The HE shrinkage procedure

perfectly corrects the EIV problem, thereby resolving the inferential issue in the baseline FE case.

Remark 3.2. In Section 3.2, we showed that under heteroskedasticity, HO yields invalid inference for β . By contrast, as we show in Proposition D.1, HE delivers an asymptotically unbiased estimator of β under the same conditions. In this sense, HE provides a more robust approach. Details and proofs are given in Appendix D.

We further establish consistency of standard errors from the regression of Y_i on $\hat{\theta}_{i,\text{HE}}$. Together with Theorem 3.3, this ensures that regression of Y_i on $\hat{\theta}_{i,\text{HE}}$ delivers automatic inference in practice: one can proceed as if $\hat{\theta}_{i,\text{HE}}$ are the true latent θ_i for the purpose of estimation and inference on β .

Building on Assumption 3.1, additional regularity conditions are required for consistency of the variance estimator.

Assumption 3.2. 1. $\mathbb{E}(u_i) = 0$, $\mathbb{E}(u_i\theta_i) = 0$. $\mathbb{E}[|\theta_i^{k_1}u_i^{k_2}|] < \infty$ for any $1 \leq k_1 \leq 8$, $1 \leq k_2 \leq 4$.

2. $\mathbb{E}[\theta_i^8] < \infty$.

Here, Assumption 3.2 strengthens the moment conditions to ensure that the law of large numbers applies to higher order terms in the variance estimator.

Theorem 3.4. *Suppose Assumption 3.1 and Assumption 3.2 hold. Then the Eicker–Huber–White estimator of standard errors in the regression of Y_i on $\hat{\theta}_{i,\text{HE}}$ is consistent:*

$$\hat{\Omega} := \frac{\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,\text{HE}} - \bar{\theta}_{\text{HE}} \right)^2 \hat{u}_i^2}{\left(\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,\text{HE}} - \bar{\theta}_{\text{HE}} \right)^2 \right)^2} \xrightarrow{p} \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2]}{(\text{Var}(\theta_i))^2},$$

where \hat{u}_i is the OLS residual:

$$\hat{u}_i := Y_i - \hat{\alpha} - \hat{\beta}\hat{\theta}_{i,\text{HE}}.$$

Taken together, Theorem 3.3 and Theorem 3.4 show that asymptotically valid confidence intervals for β can be constructed as follows. If we regress Y_i on $\hat{\theta}_{i,\text{HE}}$, the reported standard error is $\sqrt{\hat{\Omega}/n}$. The asymptotically valid confidence intervals at

level $1 - \alpha$ are given by

$$\hat{\beta}_{\text{HE}} \pm z_{1-\alpha/2} \sqrt{\hat{\Omega}/n},$$

where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution. Theorem 3.3 and Theorem 3.4 ensure that the confidence interval has asymptotically nominal coverage. Specifically:

$$\Pr \left(\beta \in [\hat{\beta}_{\text{HE}} - z_{1-\alpha/2} \sqrt{\hat{\Omega}/n}, \hat{\beta}_{\text{HE}} + z_{1-\alpha/2} \sqrt{\hat{\Omega}/n}] \right) \rightarrow 1 - \alpha.$$

In practice, efficient inference is easy to implement. One regresses Y_i on $\hat{\theta}_{i,\text{HE}}$ to obtain the OLS estimator $\hat{\beta}_{\text{HE}}$ and its standard error. The resulting confidence interval for β is asymptotically valid.

The HE estimator also serves another purpose. While the focus of this paper is on downstream inference, it is worth noting that the HE estimator has favorable properties for estimating the full vector of individual effects $\boldsymbol{\theta}$. As an estimator motivated by empirical Bayes principles, its construction improves estimation accuracy (see, e.g., Efron and Morris, 1973). By correctly modeling all sources of heterogeneity, $\hat{\boldsymbol{\theta}}_{\text{HE}} := \left(\hat{\theta}_{i,\text{HE}} \right)_{i=1}^n$ is known to achieve a lower mean squared error (MSE) for $\boldsymbol{\theta}$ than the unshrunk FE, the HO, and the common-weight estimators, a standard result in that literature.

3.4 Common-Weight Shrinkage (CW)

As a final point of comparison, we analyze the common-weight (CW) shrinkage estimator, which serves as a benchmark. This estimator, motivated by the James-Stein estimator, shrinks each unit mean \bar{X}_i towards the grand mean \bar{X} by a common factor w :

$$\hat{\theta}_{i,\text{CW}} := w \bar{X}_i + (1 - w) \bar{X},$$

where \bar{X} denotes the grand mean of the sample. The weight w can be data-dependent as in the James-Stein estimator. For instance, common-weight shrinkage is applied in Chetty et al. (2014a,b), where the weight w is chosen from the best linear predictor of θ_i given \bar{X}_i . Inference of β is then performed by regressing Y_i on $\hat{\theta}_{i,\text{CW}}$.

We now give two examples of common weight w for shrinkage.

(i) Bias Correction Shrinkage Regressing Y_i on $\hat{\theta}_{i,\text{CW}}$ leads to an OLS estimator of β given by

$$\hat{\beta}_{\text{CW}} := \frac{\widehat{\text{Cov}}(Y_i, \hat{\theta}_{i,\text{CW}})}{\widehat{\text{Var}}(\hat{\theta}_{i,\text{CW}})} = w^{-1} \underbrace{\frac{\widehat{\text{Cov}}(Y_i, \bar{X}_i)}{\widehat{\text{Var}}(\bar{X}_i)}}_{\hat{\beta}_{\text{FE}}}.$$

In effect, regressing Y_i on $\hat{\theta}_{i,\text{CW}}$ adjusts $\hat{\beta}_{\text{FE}}$ of regressing Y_i on \bar{X}_i by a factor of w^{-1} . By a proper choice of shrinkage weight $w \approx \widehat{\text{Var}}(\theta_i) / \widehat{\text{Var}}(\bar{X}_i)$, the adjustment overlaps with the bias correction method for the EIV problem. Heuristically, in that case

$$\hat{\beta}_{\text{CW}} = \frac{\widehat{\text{Var}}(\bar{X}_i)}{\widehat{\text{Var}}(\theta_i)} \cdot \frac{\widehat{\text{Cov}}(Y_i, \bar{X}_i)}{\widehat{\text{Var}}(\bar{X}_i)} = \frac{\widehat{\text{Cov}}(Y_i, \bar{X}_i)}{\widehat{\text{Var}}(\theta_i)} \approx \frac{\widehat{\text{Cov}}(Y_i, \theta_i)}{\widehat{\text{Var}}(\theta_i)}.$$

The bias correction shrinkage method with the above w thus has the equivalent property as classical bias correction in EIV problems for downstream regressions.

(ii) IV Shrinkage Another choice of w would also result in a shrinkage option that, when used as the regressor, refines the IV estimator for β . To see this, if we split the measurements for each unit i into two subsets, and treat the subset averages $\bar{X}_{i,1}$, $\bar{X}_{i,2}$ as the instrument and endogenous variable respectively, and then we construct an IV estimator:

$$\hat{\beta}_{\text{IV}} = \left(\frac{\widehat{\text{Cov}}(\bar{X}_{i,1}, \bar{X}_{i,2})}{\widehat{\text{Var}}(\bar{X}_{i,1})} \right)^{-1} \frac{\widehat{\text{Cov}}(Y_i, \bar{X}_{i,1})}{\widehat{\text{Var}}(\bar{X}_{i,1})} = \frac{\widehat{\text{Cov}}(Y_i, \bar{X}_{i,1})}{\widehat{\text{Cov}}(\bar{X}_{i,1}, \bar{X}_{i,2})} \approx \frac{\widehat{\text{Cov}}(Y_i, \theta_i)}{\widehat{\text{Var}}(\theta_i)}.$$

The IV estimator adjusts the estimator from regressing Y_i on $\bar{X}_{i,1}$ by a factor. With a proper modification of the factor, it can also adjust the estimator from regressing Y_i on \bar{X}_i , which exploits more data.

Because of its connection to them, common-weight shrinkage can perform equivalently to the bias correction method and the IV estimator. It also requires as weak assumptions as theirs. Even though it ignores the heterogeneous signal-to-noise ratio and pools every individual at a uniform ratio, common-weight shrinkage provides

valid inference on β . The following result formalizes this point for the bias correction shrinkage, with the variance estimator for θ_i the same as in $\hat{\beta}_{\text{HE}}$. It is proved in Appendix B.

Proposition 3.5. *Suppose Assumption 3.1 holds. Then we have*

$$\sqrt{n} \left(\hat{\beta}_{\text{CW}} - \beta \right) \rightarrow_d N \left(0, \frac{\mathbb{E} \left[u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2 \right]}{(\text{Var}(\theta_i))^2} \right).$$

Therefore, the CW estimator also addresses the inferential problem from the FE baseline. Its primary distinction from the HE estimator is how it addresses the problem: the CW approach applies a uniform correction factor (equivalent to bias-correction or IV), while the HE estimator models unit-level heterogeneity. Our focus on individual-weight shrinkage is motivated by its prevalence in empirical applications. The common-weight approach characterizes influential work such as Chetty et al. (2014b). By contrast, a distinct and extensive literature has focused on individualized shrinkage (e.g., Jacob and Lefgren, 2007; Kane and Staiger, 2008; Andrabi et al., 2025). Our analysis thus provides the necessary theoretical foundation for this widely used empirical strategy.

4 Simulations

4.1 Simulation Design

In the first set of simulations, we focus on the comparison between regressing on the shrinkage estimator $\hat{\theta}_{i,\text{HE}}$, the sample mean \bar{X}_i , i.e. $\hat{\theta}_{i,\text{FE}}$, and the true latent θ_i . In each of the $S = 3000$ simulations, we generate the data, run the regression of Y_i on shrunk or unshrunk regressors, and then report the 95% confidence intervals of β . Finally, for each value of β in the grid, we compute the coverage rate across all simulations, i.e., the proportion of simulations in which the value of β falls within the 95% confidence interval.

The number of measurements is fixed at $J_i = J = 20$ and the sample size is $n = 1000$. Here the ratio $\hat{\kappa} \approx 1.58$. We set the true θ_i to be drawn from the standard normal distribution, and the variance σ_i^2 drawn from $\chi^2(1)$. In the regression, we set $\alpha = 0$, $\beta = 1$, and draw u_i from a standard normal—thus homoskedastic—distribution.

In the first setting, we generate the measurement error $\epsilon_{i,j}$ from a normal distribution, with all conditions in Assumption 3.1 satisfied. In the second setting, we set the measurement error $\epsilon_{i,j}$ still having the variance σ_i^2 , but following a centered linear transform $\sigma_i \frac{\chi^2(2)-2}{2}$. Note that the ratio $\epsilon_{i,j}/\sigma_i$ satisfies the moment conditions in Assumption 3.1. As $\epsilon_{i,j}$ follows a shifted Gamma distribution, which is non-normal and asymmetric, the purpose is to show that the proposed methods perform well under non-normal measurement error (provided the regularity conditions are satisfied).

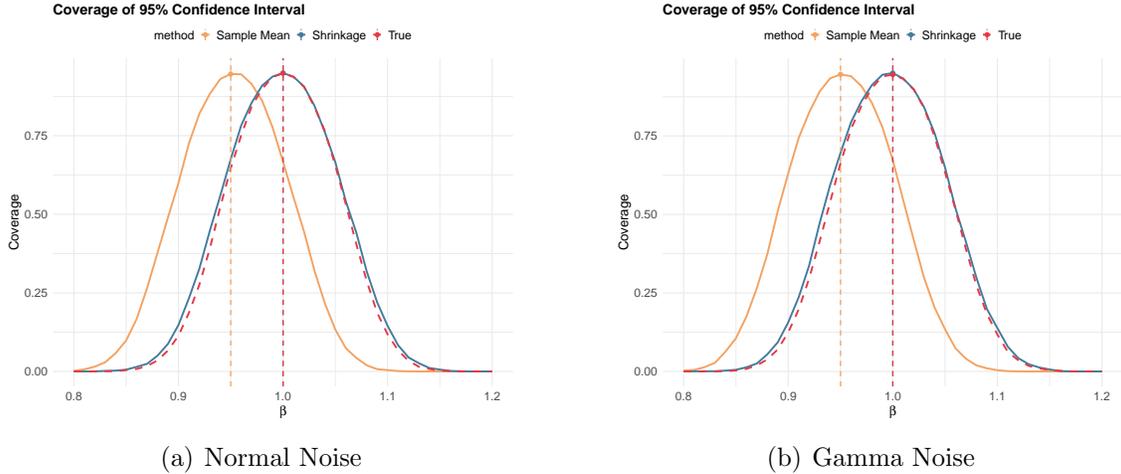
In the second set of simulations, we compare bias and coverage rates across different methods, especially on the comparison between the individual-weight method HE and CW. We report the square root scaled MSE of $\hat{\beta}$, coverage rate, bias, and MSE of $\hat{\theta}$. The design we consider lets the heteroskedastic σ_i^2 be from a uniform distribution on $\{1, 10\}$, with other configurations the same as the previous case.

In the third set of simulations, we again compute the coverage rate for each method. We check the performance with random J_i drawn from Poisson(20), and with the sample size $n = 1000$. Here $\hat{\kappa}$ is approximately 1.66, larger than 1.58 in the first set of simulations due to convexity. With other parameters unchanged, now we can introduce correlation between J_i and σ_i^2 . In the first setting, we generate J_i and σ_i^2 independently, ensuring that all conditions in Assumption 3.1 are satisfied. In the second setting, we introduce a positive correlation between J_i and σ_i^2 , reflecting possible endogenous choice of more measurements for lower precision, keeping the conditions in Assumption D.1 satisfied.

4.2 Results

The first set of coverage results is presented in Figure 1. Coverage for regressing Y_i on $\hat{\theta}_{i,\text{HE}}$ (the blue line) performs well in both normal and non-normal settings. The coverage rates are close to 95% at the true β , and the curve is always very close to the infeasible case (the dashed red line) of regressing on the true θ_i . In both cases, regressing on the sample mean \bar{X}_i suffers from attenuation bias, reflected by a shift to the left in the coverage curve (the orange line), though the spread is approximately correct, which confirms the theory developed in Proposition 3.1.

The second set of results is reported in Table 1. The results indicate that regressing Y_i on $\hat{\theta}_{i,\text{HE}}$ has a smaller MSE of β , higher coverage rate, and smaller bias. Also, $\hat{\theta}_{i,\text{HE}}$ has smaller MSE of θ than $\hat{\theta}_{i,\text{CW}}$ when the ratio $\sqrt{n}\mathbb{E}[J_i^{-1}]$ is reasonable.



Note: Each curve represents the proportion of simulations in which the value of β on the x-axis falls within the 95% confidence interval. The dashed red line represents the infeasible case of regressing on the true θ_i . The blue line represents regressing on the shrinkage estimator $\hat{\theta}_{i,\text{HE}}$ (HE). The orange line represents regressing on the sample mean \bar{X}_i (FE).

Figure 1: Coverage Rates Under Different Noise Distributions

When n and J grow large, the performance of $\hat{\theta}_{i,\text{CW}}$ and $\hat{\theta}_{i,\text{HE}}$ becomes close, but $\hat{\theta}_{i,\text{HE}}$ always dominates and is even more preferable when the sample size is relatively small. As discussed in Section 3.4, we know that $\hat{\beta}_{\text{CW}}$ is first-order asymptotically equivalent to $\hat{\beta}_{\text{HE}}$. Combined with the simulation results, we can see that $\hat{\beta}_{\text{HE}}$ can achieve better performance in finite samples.

The coverage results for random J_i are presented in Figure 2. We mainly focus on the curves for $\hat{\beta}_{\text{HE}}$ (the blue line) and $\hat{\beta}_{\text{HO}}$ (the purple line). We can see that the $\hat{\beta}_{\text{HE}}$ delivers valid coverage in both settings, close to the infeasible case (the dashed red line). Instead, $\hat{\beta}_{\text{HO}}$ only delivers valid coverage in the first setting. When J_i and σ_i^2 are correlated, the asymptotic distribution of $\hat{\beta}_{\text{HO}}$ is centered away from the true beta, reflected by a shift to the right in the coverage plot for $\hat{\beta}_{\text{HO}}$. Similar to before, regressing on the sample mean \bar{X}_i results in a leftward biased coverage curve. As discussed in Section 3.2, the asymptotic bias in $\hat{\beta}_{\text{HO}}$ is distinct from classical EIV attenuation bias, as it shifts the estimate in the opposite direction and causes amplification.

Overall, the simulation results confirm the theoretical findings in Section 3 and support Assumption 3.1 regarding the nonnormality of $\epsilon_{i,j}$. Additionally, they validate

Table 1: Method Comparison Across Scenarios

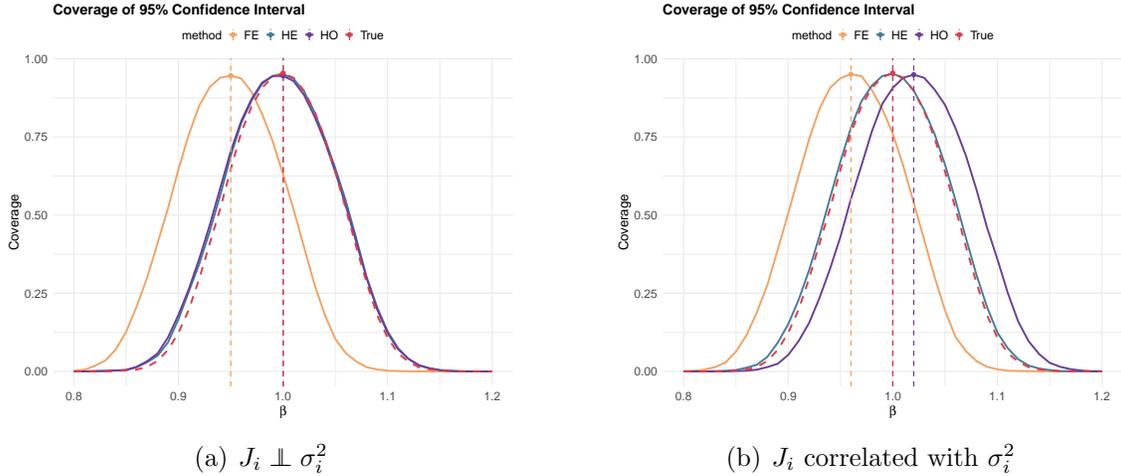
Scenario	Method	$\sqrt{n} \times \text{MSE}(\beta)$	Coverage Rate (%)	Bias	MSE(θ)
n=50, J=10 $\sqrt{n}/J \approx 0.71$	True θ_i	1.030	94.73	0.116	0
	HE	1.497	91.90	0.166	0.322
	CW	2.335	87.83	0.225	0.374
	FE	2.690	27.67	0.354	0.552
n=225, J=10 $\sqrt{n}/J \approx 1.5$	True θ_i	1.016	94.17	0.054	0
	HE	1.550	90.27	0.083	0.313
	CW	1.856	86.13	0.099	0.359
	FE	5.411	0.03	0.354	0.550
n=1000, J=20 $\sqrt{n}/J \approx 1.58$	True θ_i	1.009	95.13	0.025	0
	HE	1.298	93.13	0.033	0.195
	CW	1.392	91.33	0.035	0.216
	FE	6.930	0.00	0.217	0.275

Note: This table compares the performance of different estimation methods across various scenarios, focusing on mean squared error (MSE), coverage rate, and bias. "True θ_i " represents the benchmark case where the true values of θ_i are known. HE denotes the heteroskedastic estimator, while CW refers to the common-weight estimator. FE represents the fixed-effects estimator. The square root of scaled MSE of β is computed as $\sqrt{n} \times \text{MSE}(\beta)$. Coverage rates are reported as percentages, and bias refers to the absolute deviation of the estimated β from its true value. The MSE of θ measures the estimation error for individual effects. A larger \sqrt{n}/J indicates a higher ratio of measurement error to sampling error.

the theoretical framework in Section 3 concerning the dependence of J_i and σ_i^2 . The estimator $\hat{\beta}_{\text{HE}}$ works well in both normal and non-normal settings, and is robust to the correlation between J_i and σ_i^2 . It outperforms $\hat{\beta}_{\text{FE}}$, $\hat{\beta}_{\text{CW}}$ and $\hat{\beta}_{\text{HO}}$, and is close to the infeasible best case.

5 Empirical Application: Firm Discrimination

In this section, we illustrate the use of regression on heteroskedastic individual-weight shrinkage estimators in the context of Kline et al. (2022), which studies the extent to which large U.S. employers systemically discriminate job applicants based on race. Their study utilizes correspondence audits, where fictitious resumes with randomized



Note: Each curve represents the proportion of simulations in which the value of β on the x-axis falls within the 95% confidence interval. The dashed red line represents the infeasible case of regressing on the true θ_i . The blue line represents the method of regressing on the heteroskedastic individual-weight shrinkage estimator $\hat{\theta}_{i,HE}$ (HE). The purple line represents the regressing on the homoskedastic individual-weight shrinkage estimator $\hat{\theta}_{i,HO}$ (HO). The orange line represents regressing on the sample mean \bar{X}_i (FE).

Figure 2: Coverage Rates Under Different Dependence

racial identifiers are sent to employers to measure differences in callback rates. The racial contact gap, defined as the difference in callback probabilities between racial groups, serves as the primary latent variable, analogous to value-added in our setting.

5.1 Data

We use the panel dataset in [Kline et al. \(2022\)](#) from an experiment that sends fictitious applications to jobs posted by 108 of the largest U.S. employers. For each firm in each wave, about 25 entry-level vacancies were sampled and, for each vacancy, 8 job applications with randomly assigned characteristics were sent to the employer. Sampling was organized in 5 waves. Focusing on firms sampled in all waves yields a balanced panel of $n = 70$ firms over 5 waves. Applications were sent in pairs, one randomly assigned a distinctively White name and the other a distinctively Black name. The primary outcome is whether the employer attempted to contact the applicant within 30 days of applying. The racial contact gap is defined as the firm-level

difference between the contact rate (the ratio of the number of contacts to the number of received applications) for White and that for Black applications. We follow the model similar to [Kline et al. \(2022, p. 41\)](#), who assumes that the racial contact gap is given by

$$\bar{X}_i = \theta_i + \bar{\epsilon}_i, \quad \bar{\epsilon}_i \sim N\left(0, \frac{\sigma_i^2}{J_i}\right).$$

They justify the normality of $\bar{\epsilon}_i$ from the central limit theorem approximation with large numbers of measurements.

5.2 Estimation

In our analysis, we estimate the predictive effect of callback probability in wave t , for $t = 1, 2, 3$, on callback probability in wave $t + 2$. Given the model setup, we expect the regression coefficient to be close to 1. For each wave t , we first apply shrinkage estimation, and subsequently regress on these estimates. The shrinkage estimators constructed by [Kline et al. \(2020\)](#) are nonlinear shrinkage based on nonparametric empirical Bayes posterior means, while we focus on the performance of linear shrinkage estimators as regressors. We compare the performance of heteroskedastic individual-weight shrinkage estimator $\hat{\beta}_{\text{HE}}$ with the fixed effect estimator $\hat{\beta}_{\text{FE}}$ across the three waves. Since the discrimination gap is computed from pairs of job applications, we have $J = 100$, resulting in a small ratio of measurement error to sampling error, $\sqrt{n}/J = 0.08$.

The results presented in [Table 2](#) indicate that regression on the shrinkage estimates ($\hat{\theta}_{i,\text{HE}}$) yields coefficients closer to 1. Despite a small \sqrt{n}/J , $\hat{\beta}_{\text{FE}}$ still exhibits attenuation bias. In terms of inference, $\hat{\beta}_{\text{HE}}$ robustly rejects the null hypothesis of no predictive effect, whereas $\hat{\beta}_{\text{FE}}$ fails to reject this null hypothesis for waves 1 and 2. These findings demonstrate that $\hat{\beta}_{\text{HE}}$ improves the validity of inference for the regression coefficient.

6 Application: School Value-Added in Pakistan

In this section, we illustrate the use of regression on heteroskedastic individual-weight shrinkage estimators in the context of [Andrabi et al. \(2025\)](#) about the school value-

Table 2: Regression Results for Different Waves Predicting $t + 2$

Wave	$\hat{\beta}_{\text{FE}}$				$\hat{\beta}_{\text{HE}}$			
	β	SE	CI	p -value	β	SE	CI	p -value
Wave 1	0.150	0.100	[-0.047, 0.347]	0.140	0.987	0.385	[0.232, 1.742]	0.013
Wave 2	0.092	0.110	[-0.124, 0.308]	0.406	0.883	0.363	[0.171, 1.594]	0.017
Wave 3	0.415	0.125	[0.170, 0.659]	0.001	2.186	0.963	[0.299, 4.073]	0.026

Note: This table presents regression results for firm discrimination in wave t predicting firm discrimination in wave $t + 2$. Columns labeled \bar{X}_i represent regressions using the raw sample mean, while columns labeled $\hat{\theta}_{i,\text{HE}}$ correspond to regressions using the shrinkage estimates (HE). The coefficients (β) indicate the estimated effect of firm discrimination in wave t on wave $t + 2$. The standard error (SE), 95% confidence interval (CI), and p -value are also reported, based on the Eicker–Huber–White variance estimator.

added in Pakistan. Their study investigates the relation between private school fees and school value-added (SVA), and finds evidence that parents are able to identify and reward school quality.

6.1 Data and Empirical Setting

The analysis utilizes the LEAPS project dataset, a rich longitudinal panel of student test scores from rural Punjab, Pakistan. This setting is characterized by the rapid emergence of a private school market, making the estimation of school quality (value-added) and its perceived return (school fees) a central question of interest. We use the school-year level sample from their analysis, which consists of $n = 1158$ observations. We also have $\mathbb{E}_n [J_i^{-1}] = 0.120$. Therefore $\hat{\kappa} = \sqrt{n}\mathbb{E}_n [J_i^{-1}] \approx 4.08$.

6.2 Estimation and Results

We replicate and extend the primary downstream analysis in [Andrabi et al. \(2025\)](#), which is a regression of private school fees on estimated SVA. The latent variable θ_i represents the true SVA of a school, and the downstream regression investigates whether schools with higher SVA command higher fees.

We compare two individualized shrinkage estimators. The first is HO, as used in [Andrabi et al. \(2025\)](#). The second is HE, which is fully individualized by allowing

Table 3: Regression Results of Private School Fees on School Value-Added

Controls	$\hat{\beta}_{\text{HO}}$				$\hat{\beta}_{\text{HE}}$			
	β	SE	CI	p -value	β	SE	CI	p -value
No	991.37	320.96	[362.69, 1,620.06]	0.003	1054.13	298.15	[469.01, 1,639.25]	0.001
Yes	719.86	280.88	[169.34, 1,270.38]	0.012	793.01	264.05	[275.48, 1,310.55]	0.004

Note: This table reports coefficients from regressions of private school fees on estimated school value-added using two estimators. Columns labeled $\hat{\beta}_{\text{HO}}$ use the homoskedastic individual-weight estimator following [Andrabi et al. \(2025\)](#), while columns labeled $\hat{\beta}_{\text{HE}}$ use the heteroskedastic individual-weight estimator. The rows correspond to specifications without controls and with controls for mean mother education, mean father education, and mean household asset index. For each specification, the table reports the coefficient (β), standard error (SE), 95% confidence interval (CI), and associated p -value. Standard errors are clustered at the village level.

for heterogeneity in both J_i and the noise variance σ_i^2 .

The results are presented in Table 3 (full sample) and Table 4 (a selected sample restricting $20 \leq J_i \leq 80$, as a robustness check under less heterogeneity). The tables compare the HE estimator (right column) and the HO estimator (left column) in both bivariate specifications and specifications including household-level controls (parental education and asset index).

Across all specifications, our results confirm the central economic finding of [Andrabi et al. \(2025\)](#): SVA is a large, positive, and statistically significant predictor of private school fees. In the full-sample specification with controls (Table 3, Row 2), the HE point estimate is 793.01 (significant at the 1% level), which is comparable to the HO estimate. This comparison provides a valuable robustness check. Since one cannot know a priori if the assumptions required for valid inference with $\hat{\beta}_{\text{HO}}$ hold, the stability of the coefficients under the more general $\hat{\beta}_{\text{HE}}$ estimator strengthens the credibility of the original economic conclusion.

7 Conclusion

This paper investigates the inferential properties of using individualized shrinkage estimators as covariates in a regression, a widely used but not fully understood em-

Table 4: Regression Results of Private School Fees on School Value-Added
(Selected Sample)

Controls	$\hat{\beta}_{\text{HO}}$				$\hat{\beta}_{\text{HE}}$			
	β	SE	CI	p -value	β	SE	CI	p -value
No	1569.70	451.27	[685.22, 2454.18]	0.000	1622.13	454.32	[733.66, 2510.60]	0.000
Yes	1320.18	384.74	[566.90, 2073.46]	0.000	1358.51	391.16	[592.06, 2124.96]	0.000

Note: This table reports coefficients from regressions of private school fees on estimated school value-added using two estimators. The sample is restricted to schools with between 20 and 80 students. Columns labeled $\hat{\beta}_{\text{HO}}$ use the homoskedastic individual-weight estimator following [Andrabi et al. \(2025\)](#), while columns labeled $\hat{\beta}_{\text{HE}}$ use the heteroskedastic individual-weight estimator. The rows correspond to specifications without controls and with controls for mean mother education, mean father education, and mean household asset index. For each specification, the table reports the coefficient (β), standard error (SE), 95% confidence interval (CI), and associated p -value. Standard errors are clustered at the village level.

pirical practice. We formalize the conditions under which this approach yields valid inference in the downstream regression model. Our central finding is that a correctly specified and fully individualized shrinkage estimator yields an asymptotically efficient estimator of the downstream regression coefficient. Crucially, we show that conventional OLS inference is asymptotically valid, justifying standard empirical practice. This result stands in contrast to the biased inference from regressing on the unshrunk fixed effect estimates (individual means). Similarly, the simpler individual-weight (HO) estimators which are widely used in practice can fail to deliver valid inference when noise variance is correlated with the number of measurements. Our simulations confirm the predictions of the theory, including the finding that regression on individualized shrinkage estimators that do not properly account for noise heteroskedasticity suffers from amplification bias. We apply our method to data on firm discrimination and school value-added and show that it improves the estimation and inference.

Our analysis provides a formal bridge between the empirical Bayes literature, where shrinkage estimators were developed to improve estimation accuracy of individual effects, and the common empirical practice of using these estimates for downstream inference. The key takeaway for practitioners is that while the plug-in approach can be valid and efficient, its robustness depends critically on the specification of the shrinkage estimator. For applied researchers, our results provide a clear the-

oretical foundation and a practical guide for obtaining valid inference when using shrinkage estimates as regressors in linear models. Extending this method to nonlinear settings is a natural yet nontrivial direction, which we are studying in ongoing work.

A Proofs of Main Results

A.1 Proofs for FE

Proof of Proposition 3.1. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{\text{FE}}$ as $\hat{\beta}$, and $\text{Var}(\theta_i)$ as V .

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 \right)^{-1} \\ &\quad \left(\beta \frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{X}_i - \bar{X})(\theta_i - \bar{\theta}) - \beta \frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{X}_i - \bar{X})(u_i - \bar{u}) \right) \\ &= \frac{\beta \sqrt{n} T_{1,n} - \beta \sqrt{n} T_{2,n} + \sqrt{n} T_{3,n}}{T_{2,n}}. \end{aligned}$$

Firstly, from Lemma B.1 and Lemma B.3, we have for the denominator,

$$\begin{aligned} T_{2,n} &= \hat{V} + \frac{n-1}{n^2} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 \\ &= \hat{V} + O_p(n^{-1/2}) \\ &= V + O_p(n^{-1/2}). \end{aligned}$$

For the numerator terms, by properties from Lemma B.1,

$$\begin{aligned} \sqrt{n} T_{1,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i (\theta_i - \mathbb{E}[\theta_i]) \\ &\quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{\epsilon} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + o_p(1), \end{aligned}$$

$$\begin{aligned}
\sqrt{n}T_{3,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i u_i \\
&\quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{u} - \sqrt{n} \bar{\epsilon} \bar{u} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + o_p(1).
\end{aligned}$$

Combined with the proof of Lemma B.3, we have

$$\begin{aligned}
\sqrt{n}T_{2,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2] - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \sqrt{n} \bar{\epsilon}^2 + \frac{2}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i \\
&\quad - 2\sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{\epsilon} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n [(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2] + o_p(1).
\end{aligned}$$

Therefore, the numerator is

$$\begin{aligned}
&\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n} \\
&= -\beta \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[(\theta_i - \mathbb{E}[\theta_i]) u_i + \beta \frac{1}{J_i} \sigma_i^2 - \beta \bar{\epsilon}_i^2 \right] - \beta \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{J_i} \sigma_i^2 + o_p(1) \\
&:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i - \beta \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{J_i} \sigma_i^2 + o_p(1).
\end{aligned}$$

Here for the second term, by Chebyshev's inequality, for any $s > 0$ we have

$$\Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \sigma_i^2 - \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \right] \right| > s \right) \leq \frac{\mathbb{E} \left[\frac{1}{J_i^2} \right] \mathbb{E} [\sigma_i^4]}{s^2} \rightarrow 0,$$

and also,

$$\sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \right] = \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^2] \rightarrow \kappa \mathbb{E} [\sigma_i^2].$$

Therefore, the second term is $-\kappa\beta\mathbb{E}[\sigma_i^2] + o_p(1)$.

For ξ_i , since

$$\mathbb{E} [\xi_i] = 0,$$

$$\begin{aligned} \mathbb{E} [\xi_i^2] &= \mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2] + \beta^2 \mathbb{E} \left[\left(\frac{1}{J_i} \sigma_i^2 \right)^2 \right] + \beta^2 \mathbb{E} [\bar{\epsilon}_i^4] - 2\beta^2 \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \bar{\epsilon}_i^2 \right] \\ &= \mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2] + o(1) - 2\beta^2 \mathbb{E} \left[\frac{1}{J_i^2} \sigma_i^4 \right] \\ &= \mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2] + o(1), \end{aligned} \tag{4}$$

where the last line follows from Lemma B.1.

Therefore, for $(n\mathbb{E} [\xi_i^2])^{-1/2} \sum_{i=1}^n \xi_i$ in the triangular array, by the Lindeberg-Feller theorem (See Ferguson (2017) p.27), because the Lindeberg condition holds below

$$\begin{aligned} \frac{1}{n\mathbb{E} [\xi_i^2]} \sum_{i=1}^n \mathbb{E} \left\{ \xi_i^2 \mathbb{1} \left(|\xi_i| > s \sqrt{n\mathbb{E} [\xi_i^2]} \right) \right\} &= \frac{1}{\mathbb{E} [\xi_i^2]} \mathbb{E} \left\{ \xi_i^2 \mathbb{1} \left(|\xi_i| > s \sqrt{n\mathbb{E} [\xi_i^2]} \right) \right\} \\ &\rightarrow 0, \forall s > 0, \end{aligned}$$

which is derived from (4) and the dominated convergence theorem, then we have

$$\frac{1}{\sqrt{n\mathbb{E} [\xi_i^2]}} \sum_{i=1}^n \xi_i \rightarrow_d N(0, 1).$$

By (4) and Slutsky's theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{J_i} \sigma_i^2 \rightarrow_d N \left(-\kappa\beta \mathbb{E} [\sigma_i^2], \mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2] \right).$$

Then combined with the denominator, by Slutsky's theorem,

$$\sqrt{n} (\hat{\beta} - \beta) = \frac{\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n}}{T_{2,n}} \rightarrow_d N \left(-\kappa\beta \frac{\mathbb{E} [\sigma_i^2]}{V}, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{V^2} \right).$$

□

A.2 Proofs for Asymptotic Normality and Inference

Proof of Lemma 3.1. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{c,\text{HE}}$ as $\hat{\beta}_c$, $\hat{\theta}_{i,c,\text{HE}}$ as $\hat{\theta}_{i,c}$, and $\text{Var}(\theta_i)$ as V .

We have

$$\begin{aligned}
& \sqrt{n} \left(\hat{\beta}_c - \beta \right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 \right)^{-1} \\
& \left(\beta \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right) (\theta_i - \bar{\theta}) - \beta \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right) (u_i - \bar{u}) \right) \\
&:= \frac{\beta \sqrt{n} T_{1,n} - \beta \sqrt{n} T_{2,n} + \sqrt{n} T_{3,n}}{T_{2,n}}.
\end{aligned}$$

Respectively, by properties from Lemma B.1,

$$\begin{aligned}
\sqrt{n} T_{1,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i \left[(\theta_i - \bar{\theta})^2 + (\bar{\epsilon}_i - \bar{\epsilon}) (\theta_i - \bar{\theta}) \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i (\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i \bar{\epsilon}_i (\theta_i - \mathbb{E}[\theta_i]) \\
& \quad - 2\sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i (\theta_i - \mathbb{E}[\theta_i]) + \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i])^2 \frac{1}{n} \sum_{i=1}^n c_i \\
& \quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i \bar{\epsilon}_i - \sqrt{n} \bar{\epsilon} \frac{1}{n} \sum_{i=1}^n c_i (\theta_i - \mathbb{E}[\theta_i]) + \sqrt{n} \bar{\epsilon} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i (\theta_i - \mathbb{E}[\theta_i])^2 + o_p(1),
\end{aligned}$$

$$\begin{aligned}
T_{2,n} &= \frac{1}{n} \sum_{i=1}^n (c_i (\theta_i - \bar{\theta}) + c_i (\bar{\epsilon}_i - \bar{\epsilon}))^2 + \left[\frac{1}{n} \sum_{i=1}^n (c_i (\theta_i - \bar{\theta} + \bar{\epsilon}_i - \bar{\epsilon})) \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n c_i^2 \left[(\theta_i - \bar{\theta})^2 + (\bar{\epsilon}_i - \bar{\epsilon})^2 + 2 (\theta_i - \bar{\theta}) (\bar{\epsilon}_i - \bar{\epsilon}) \right] \\
&\quad + \left[\frac{1}{n} \sum_{i=1}^n c_i \theta_i - \bar{\theta} \bar{c} + \frac{1}{n} \sum_{i=1}^n c_i \bar{\epsilon}_i - \bar{\epsilon} \bar{c} \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n c_i^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2 + 2 (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i \right] \\
&\quad - 2 (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i^2 (\theta_i - \mathbb{E}[\theta_i]) + (\bar{\theta} - \mathbb{E}[\theta_i])^2 \frac{1}{n} \sum_{i=1}^n c_i^2 \\
&\quad - 2 \bar{\epsilon} \frac{1}{n} \sum_{i=1}^n c_i^2 \bar{\epsilon}_i + \bar{\epsilon}^2 \frac{1}{n} \sum_{i=1}^n c_i^2 \\
&\quad - 2 (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i^2 \bar{\epsilon}_i - 2 \bar{\epsilon} \frac{1}{n} \sum_{i=1}^n c_i^2 (\theta_i - \mathbb{E}[\theta_i]) + 2 \bar{\epsilon} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i^2 \\
&\quad + \left[\mathbb{E}[\theta_i] + O_p(n^{-1/2}) - (\mathbb{E}[\theta_i] + O_p(n^{-1/2})) (1 + O_p(n^{-1/2})) + o_p(n^{-1/2}) \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^n c_i^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2 \right] + o_p(n^{-1/2}) \\
&\quad - 2 O_p(n^{-1/2}) O_p(n^{-1/2}) + O_p(n^{-1}) [1 + O_p(n^{-1/2})] \\
&\quad - 2 o_p(n^{-1/2}) o_p(n^{-1/2}) + o_p(n^{-1}) [1 + O_p(n^{-1/2})] \\
&\quad - 2 O_p(n^{-1/2}) o_p(n^{-1/2}) - 2 o_p(n^{-1/2}) O_p(n^{-1/2}) + 2 o_p(n^{-1/2}) O_p(n^{-1/2}) [1 + O_p(n^{-1/2})] \\
&\quad + O_p(n^{-1}) \\
&= \frac{1}{n} \sum_{i=1}^n c_i^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2 \right] + o_p(n^{-1/2}).
\end{aligned}$$

$$\begin{aligned}
\sqrt{n}T_{3,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i [(\theta_i - \bar{\theta})(u_i - \bar{u}) + (\bar{\epsilon}_i - \bar{\epsilon})(u_i - \bar{u})] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i [(\theta_i - \mathbb{E}[\theta_i]) u_i] + \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i \bar{\epsilon}_i u_i \\
&\quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i u_i - \sqrt{n} \bar{u} \frac{1}{n} \sum_{i=1}^n c_i (\theta_i - \mathbb{E}[\theta_i]) + \sqrt{n} \bar{u} (\bar{\theta} - \mathbb{E}[\theta_i]) \frac{1}{n} \sum_{i=1}^n c_i \\
&\quad - \sqrt{n} \bar{\epsilon} \frac{1}{n} \sum_{i=1}^n c_i u_i - \sqrt{n} \bar{u} \frac{1}{n} \sum_{i=1}^n c_i \bar{\epsilon}_i + \sqrt{n} \bar{u} \bar{\epsilon} \frac{1}{n} \sum_{i=1}^n c_i \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n c_i [(\theta_i - \mathbb{E}[\theta_i]) u_i] + o_p(1).
\end{aligned}$$

Therefore, for the denominator we have

$$\begin{aligned}
T_{2,n} &= \frac{1}{n} \sum_{i=1}^n c_i^2 [(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2] + o_p(n^{-1/2}) \\
&= \frac{1}{n} \sum_{i=1}^n c_i^2 (\theta_i - \mathbb{E}[\theta_i])^2 + O_p(n^{-1/2}) \\
&= V + o_p(1).
\end{aligned}$$

Also, combined with Lemma B.2, the numerator is

$$\begin{aligned}
&\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n [\beta c_i (\theta_i - \mathbb{E}[\theta_i])^2 - \beta c_i^2 [(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2] + c_i (\theta_i - \mathbb{E}[\theta_i]) u_i] + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\beta c_i (\theta_i - \mathbb{E}[\theta_i])^2 - \beta c_i^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{J_i} \hat{\sigma}_i^2 \right] + c_i (\theta_i - \mathbb{E}[\theta_i]) u_i \right] + o_p(1) \\
&:= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i + o_p(1).
\end{aligned}$$

For ξ_i , since

$$\begin{aligned}
\mathbb{E} [\xi_i] &= \beta \mathbb{E} [c_i (\theta_i - \mathbb{E} [\theta_i])^2] - \beta \mathbb{E} \left[c_i^2 \left((\theta_i - \mathbb{E} [\theta_i])^2 + \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \\
&= \beta V \mathbb{E} (c_i) - \beta \mathbb{E} \left[c_i^2 \left(V + \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \\
&= \beta V \mathbb{E} (c_i) - \beta \mathbb{E} [V c_i] \\
&= 0, \\
\mathbb{E} [\xi_i^2] &= \beta^2 \mathbb{E} [(c_i - c_i^2)^2] \mathbb{E} [(\theta_i - \mathbb{E} (\theta_i))^4] + \beta^2 \mathbb{E} \left[\frac{1}{J_i^2} c_i^4 \hat{\sigma}_i^4 \right] \\
&\quad + \mathbb{E} [c_i^2] \mathbb{E} [u_i^2 (\theta - \mathbb{E} [\theta_i])^2] \\
&\quad - 2\beta^2 V \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 c_i^2 (c_i - c_i^2) \right] \\
&= o(1) + \beta^2 \mathbb{E} \left[c_i^4 \left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right] \\
&\quad + \mathbb{E} [c_i^2] \mathbb{E} [u_i^2 (\theta - \mathbb{E} [\theta_i])^2] - 2\beta^2 V \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 c_i^2 (c_i - c_i^2) \right] \\
&= \mathbb{E} [u_i^2 (\theta - \mathbb{E} [\theta_i])^2] + o(1), \tag{5}
\end{aligned}$$

where the last line follows from Lemma B.1 and also $0 < c_i < 1$.

Therefore, for $(n\mathbb{E} [\xi_i^2])^{-1/2} \sum_{i=1}^n \xi_i$ in the triangular array, by the Lindeberg-Feller theorem (See Ferguson (2017) p.27), because the Lindeberg condition holds below

$$\begin{aligned}
\frac{1}{n\mathbb{E} [\xi_i^2]} \sum_{i=1}^n \mathbb{E} \left\{ \xi_i^2 \mathbb{1} \left(|\xi_i| > s \sqrt{n\mathbb{E} [\xi_i^2]} \right) \right\} &= \frac{1}{\mathbb{E} [\xi_i^2]} \mathbb{E} \left\{ \xi_i^2 \mathbb{1} \left(|\xi_i| > s \sqrt{n\mathbb{E} [\xi_i^2]} \right) \right\} \\
&\rightarrow 0, \quad \forall s > 0,
\end{aligned}$$

which is derived from (5) and the dominated convergence theorem, then we have

$$\frac{1}{\sqrt{n\mathbb{E} [\xi_i^2]}} \sum_{i=1}^n \xi_i \rightarrow_d N(0, 1).$$

By (5) and Slutsky's theorem,

$$\frac{1}{\sqrt{n\sigma_u^2 V}} \sum_{i=1}^n \xi_i \rightarrow_d N(0, 1).$$

Therefore,

$$\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n} \rightarrow_d N\left(0, \mathbb{E}\left[u_i^2(\theta - \mathbb{E}[\theta_i])^2\right]\right).$$

Then combined with the denominator, by Slutsky's theorem,

$$\sqrt{n}\left(\hat{\beta}_c - \beta\right) = \frac{\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n}}{T_{2,n}} \rightarrow_d N\left(0, \frac{\mathbb{E}\left[u_i^2(\theta - \mathbb{E}[\theta_i])^2\right]}{V^2}\right).$$

□

Proof of Theorem 3.3. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{c,\text{HE}}$ as $\hat{\beta}_c$, $\hat{\beta}_{\text{HE}}$ as $\hat{\beta}$, $\hat{\theta}_{i,c,\text{HE}}$ as $\hat{\theta}_{i,c}$, $\hat{\theta}_{i,\text{HE}}$ as $\hat{\theta}_i$ and $\text{Var}(\theta_i)$ as V . We then prove the asymptotics by showing

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \rightarrow_p \sqrt{n}\left(\hat{\beta}_c - \beta\right).$$

By taking the difference, it's equivalent to

$$\sqrt{n}\hat{\beta} - \sqrt{n}\hat{\beta}_c = o_p(1).$$

It suffices to show for the numerator part

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\theta}_i - \bar{\theta}\right) (Y_i - \bar{Y}) \rightarrow_p \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c\right) (Y_i - \bar{Y}),$$

and then show for the denominator part

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_i - \bar{\theta}\right)^2 \rightarrow_p \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c\right)^2.$$

First we show the numerator part. Since $\hat{\theta}_i$ and $\hat{\theta}_{i,c}$ can be viewed as function values

given $t = \hat{V}$ and $t = V$ for the function of t :

$$\bar{X} + \frac{t}{\frac{1}{J_i} \hat{\sigma}_i^2 + t} (\bar{X}_i - \bar{X}),$$

the first order derivative of which is denoted as

$$A_i(t) := \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + t\right)^2} (\bar{X}_i - \bar{X}),$$

for simplicity of notations, we denote the whole numerator part as a function $f_n(t)$, and what we want to show is

$$f_n(\hat{V}) - f_n(V) = o_p(1).$$

By the mean value expansion theorem, for \tilde{V} such that $|\tilde{V} - V| \leq |\hat{V} - V|$,

$$\begin{aligned} \left| f_n(\hat{V}) - f_n(V) \right| &= |\hat{V} - V| \left| f_n'(\tilde{V}) \right| \\ &\leq \sqrt{n} |\hat{V} - V| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + \tilde{V}\right)^2} |(\bar{X}_i - \bar{X})(Y_i - \bar{Y})| \end{aligned}$$

For any $s > 0$, there exists $\delta > 0$ such that $V - \delta > 0$, and we have

$$\begin{aligned}
& \Pr \left(\left| f_n(\hat{V}) - f_n(V) \right| > s \right) \\
& \leq \Pr \left(\left| f_n(\hat{V}) - f_n(V) \right| > s, \left| \hat{V} - V \right| \leq \delta \right) + \Pr \left(\left| \hat{V} - V \right| > \delta \right) \\
& \leq \Pr \left(\sqrt{n} \left| \hat{V} - V \right| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + \hat{V} \right)^2} \left| (\bar{X}_i - \bar{X})(Y_i - \bar{Y}) \right| > s, \left| \hat{V} - V \right| \leq \delta \right) \\
& \quad + \Pr \left(\left| \hat{V} - V \right| > \delta \right) \\
& \leq \Pr \left(\sqrt{n} \left| \hat{V} - V \right| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \left| (\bar{X}_i - \bar{X})(Y_i - \bar{Y}) \right| > s, \left| \hat{V} - V \right| \leq \delta \right) \\
& \quad + \Pr \left(\left| \hat{V} - V \right| > \delta \right) \\
& \leq \Pr \left(\sqrt{n} \left| \hat{V} - V \right| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \left| (\bar{X}_i - \bar{X})(Y_i - \bar{Y}) \right| > s \right) \\
& \quad + \Pr \left(\left| \hat{V} - V \right| > \delta \right)
\end{aligned}$$

Then we only need to show

$$\frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \left| (\bar{X}_i - \bar{X})(Y_i - \bar{Y}) \right| = o_p(1).$$

This can be proved if for any integers $k_1, k_2 \in \{0, 1\}$,

$$\frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \left| \bar{X}_i \right|^{k_1} \left| Y_i \right|^{k_2} = o_p(1).$$

For any $s > 0$, by Chebyshev's inequality,

$$\begin{aligned}
& \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |\bar{X}_i|^{k_1} |Y_i|^{k_2} - \mathbb{E} \left[\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |\bar{X}_i|^{k_1} |Y_i|^{k_2} \right] \right| > s \right) \\
& \leq \frac{\mathbb{E} \left[\left(\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} \right)^2 |\bar{X}_i|^{2k_1} |Y_i|^{2k_2} \right]}{ns^2} \\
& \leq \frac{\frac{1}{(V-\delta)^2} \mathbb{E} \left[|\bar{X}_i|^{2k_1} (\alpha + \beta\theta_i + u_i)^{2k_2} \right]}{ns^2} \rightarrow 0.
\end{aligned}$$

where the last limit follows from the independence and finite moments assumptions, as well as the fact from Lemma B.1 that $\mathbb{E} \left[|\bar{X}_i|^{2k_1} \right] \rightarrow \mathbb{E} \left[|\theta_i|^{2k_1} \right]$. Also, by the Cauchy-Schwarz inequality, from the property of $\left(\frac{1}{J_i} \hat{\sigma}_i^2\right)^2$ in Lemma B.1, we also have

$$\begin{aligned}
\mathbb{E} \left[\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |\bar{X}_i|^{k_1} |Y_i|^{k_2} \right] &= \mathbb{E} \left[\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} \right] \mathbb{E} \left[|\bar{X}_i|^{k_1} |Y_i|^{k_2} \right] \\
&\leq \frac{1}{(V-\delta)^2} \sqrt{\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2\right)^2 \right] \mathbb{E} \left[|\bar{X}_i|^{2k_1} |Y_i|^{2k_2} \right]} \rightarrow 0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |(\bar{X}_i - \bar{X})(Y_i - \bar{Y})| \\
& \leq \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |\bar{X}_i Y_i| + |\bar{X}| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |Y_i| \\
& \quad + |\bar{Y}| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} |\bar{X}_i| + |\bar{X}| |\bar{Y}| \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta\right)^2} \\
& = o_p(1) + O_p(1) \cdot o_p(1) + O_p(1) \cdot o_p(1) + O_p(1) \cdot O_p(1) \cdot o_p(1) = o_p(1).
\end{aligned}$$

Then the numerator part is shown.

Lastly, we show for the denominator part

$$\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \rightarrow_p \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,c} - \bar{\theta}_c)^2.$$

Similarly, for simplicity of notations, we denote the whole denominator part as a function $g_n(t)$. Similarly, we have

$$g_n(\hat{V}) - g_n(V) = (\hat{V} - V) g'_n(\tilde{V}).$$

Here with a slight abuse of notation, \tilde{V} satisfies $|\tilde{V} - V| \leq |\hat{V} - V|$. We finish the proof by showing

$$g'_n(\tilde{V}) = O_p(1).$$

Let

$$D_i := \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + \tilde{V}\right)^2}.$$

$$G_i := \frac{\tilde{V}}{\frac{1}{J_i} \hat{\sigma}_i^2 + \tilde{V}}.$$

$$\begin{aligned} g'_n(\tilde{V}) &= \frac{2}{n} \sum_{i=1}^n \left[G_i (\bar{X}_i - \bar{X}) - \frac{1}{n} \sum_{k=1}^n G_k (\bar{X}_k - \bar{X}) \right] D_i (\bar{X}_i - \bar{X}) \\ &= \frac{2}{n} \sum_{i=1}^n G_i D_i (\bar{X}_i - \bar{X})^2 - 2 \left[\frac{1}{n} \sum_{i=1}^n G_i (\bar{X}_i - \bar{X}) \right] \left[\frac{1}{n} \sum_{i=1}^n D_i (\bar{X}_i - \bar{X}) \right]. \end{aligned}$$

Similarly, to show it's $O_p(1)$, it suffices to show for $k_1, k_2 \in \{0, 1\}$, $k_3 \in \{0, 1, 2\}$,

$$\frac{1}{n} \sum_{i=1}^n G_i^{k_1} D_i^{k_2} \bar{X}_i^{k_3} = O_p(1).$$

For any $M > 0$, there exists $0 < \delta < V$, and

$$\begin{aligned}
& \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n G_i^{k_1} D_i^{k_2} \bar{X}_i^{k_3} \right| > M \right) \\
& \leq \Pr \left(\frac{1}{n} \sum_{i=1}^n G_i^{k_1} D_i^{k_2} |\bar{X}_i|^{k_3} > M, |\tilde{V} - V| \leq \delta \right) + \Pr \left(|\tilde{V} - V| > \delta \right) \\
& \leq \Pr \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{V - \delta}{\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta} \right)^{k_1} \left(\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \right)^{k_2} |\bar{X}_i|^{k_3} > M \right) + \Pr \left(|\tilde{V} - V| > \delta \right) \\
& \leq \Pr \left(\frac{1}{(V - \delta)^{k_2}} \frac{1}{n} \sum_{i=1}^n |\bar{X}_i|^{k_3} > M \right) + \Pr \left(|\tilde{V} - V| > \delta \right).
\end{aligned}$$

Because

$$\frac{1}{n} \sum_{i=1}^n |\bar{X}_i|^{k_3} = O_p(1),$$

therefore it is proved. \square

Proof of Theorem 3.4. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{\text{HE}}$ as $\hat{\beta}$, $\hat{\theta}_{i,\text{HE}}$ as $\hat{\theta}_i$, and $\text{Var}(\theta_i)$ as V . For the numerator, we have

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \hat{u}_i^2 \\
& = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \left(Y_i - \bar{Y} - \hat{\beta} (\hat{\theta}_i - \bar{\theta}) \right)^2 \\
& = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 (Y_i - \bar{Y})^2 - 2\hat{\beta} \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^3 (Y_i - \bar{Y}) \\
& \quad + \hat{\beta}^2 \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^4.
\end{aligned}$$

Then by Lemma C.1, in order to show

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_i - \bar{\theta} \right)^2 \hat{u}_i^2 \rightarrow \mathbb{E} [(\theta_i - \mathbb{E}[\theta_i])^2 u_i^2],$$

it suffices to show for integers $2 \leq k \leq 4$,

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_i - \bar{\theta} \right)^k (Y_i - \bar{Y})^{4-k} \rightarrow_p \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^k (Y_i - \bar{Y})^{4-k}.$$

Because of the proof shown in Theorem 3.3, the above can be shown if we have $2 \leq k \leq 4$,

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_i - \bar{\theta} \right)^k Y_i^{4-k} \rightarrow_p \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^k Y_i^{4-k}$$

For simplicity of notations, we denote the left hand side as the function value of $f_n(t)$ at $t = \hat{V}$, and the right hand side as the function value at $t = V$. Then we have

$$f_n(\hat{V}) - f_n(V) = (\hat{V} - V) f'_n(\tilde{V}).$$

Here \tilde{V} satisfies $|\tilde{V} - V| \leq |\hat{V} - V|$. Then we only need to show

$$f'_n(\tilde{V}) = O_p(1).$$

Let

$$D_i := \frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + \tilde{V} \right)^2}.$$

$$G_i := \frac{\tilde{V}}{\frac{1}{J_i} \hat{\sigma}_i^2 + \tilde{V}}.$$

$$\begin{aligned}
& f'_n(\tilde{V}) \\
&= \frac{k}{n} \sum_{i=1}^n \left[G_i(\bar{X}_i - \bar{X}) - \frac{1}{n} \sum_{k=1}^n G_k(\bar{X}_k - \bar{X}) \right]^{k-1} \left(D_i(\bar{X}_i - \bar{X}) - \frac{1}{n} \sum_{k=1}^n D_k(\bar{X}_k - \bar{X}) \right) Y_i^{4-k}
\end{aligned}$$

It suffices to show for $k_1 \in \{1, 2, 3\}$, $k_2 \in \{0, 1\}$, $k_3 \in 1, 2, 3$,

$$\frac{1}{n} \sum_{i=1}^n G_i^{k_1} D_i^{k_2} \bar{X}_i^{k_3} Y_i^{4-k} = O_p(1).$$

For any $M > 0$, there exists $0 < \delta < V$, and

$$\begin{aligned}
& \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n G_i^{k_1} D_i^{k_2} \bar{X}_i^{k_3} Y_i^{4-k} \right| > M \right) \\
& \leq \Pr \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{V - \delta}{\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta} \right)^{k_1} \left(\frac{\frac{1}{J_i} \hat{\sigma}_i^2}{\left(\frac{1}{J_i} \hat{\sigma}_i^2 + V - \delta \right)^2} \right)^{k_2} |\bar{X}_i|^{k_3} Y_i^{4-k} > M \right) + \Pr \left(|\tilde{V} - V| > \delta \right). \\
& \leq \Pr \left(\frac{1}{(V - \delta)^{k_2}} \frac{1}{n} \sum_{i=1}^n |\bar{X}_i|^{k_3} Y_i^{4-k} > M \right) + \Pr \left(|\tilde{V} - V| > \delta \right).
\end{aligned}$$

Because

$$\frac{1}{n} \sum_{i=1}^n |\bar{X}_i|^{k_3} Y_i^{4-k} = O_p(1),$$

therefore it is proved.

For the denominator, from the proof of Theorem 3.3 and Lemma 3.1, we have

$$\left(\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \right) \rightarrow_p V^2.$$

□

References

- Abdulkadirođlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020). Do parents value school effectiveness? *American Economic Review* 110(5), 1502–1539.
- Andrabi, T., N. Bau, J. Das, and A. I. Khwaja (2025). Heterogeneity in school value added and the private premium. *American Economic Review* 115(1), 147–182.
- Angelova, J. A. (2012). On moments of sample mean and variance. *Int. J. Pure Appl. Math* 79(1), 67–85.
- Angelova, V., W. Dobbie, and C. S. Yang (2025). Algorithmic recommendations and human discretion. *Review of Economic Studies*, rdaf084.
- Battaglia, L., T. Christensen, S. Hansen, and S. Sacher (2024). Inference for regression with variables generated from unstructured data. *arXiv preprint arXiv:2402.15585*.
- Bau, N. and J. Das (2020). Teacher value added in a low-income country. *American Economic Journal: Economic Policy* 12(1), 62–96.
- Biasi, B. and H. Sarsons (2022). Flexible wages, bargaining, and the gender gap. *The Quarterly Journal of Economics* 137(1), 215–266.
- Bonhomme, S. and A. Denis (2024). Estimating heterogeneous effects: applications to labor economics. *arXiv preprint arXiv:2404.01495*.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of econometrics* 34(3), 305–334.
- Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016). Health care exceptionalism? performance and allocation in the us health care sector. *American Economic Review* 106(8), 2110–2144.
- Chang, S.-K., S.-C. Huang, Y.-C. Chen, and S.-H. Liao (2024, February). Post empirical bayes regression. Technical report, National Taiwan University and University of Chicago. Working Paper.
- Chen, J. (2023). Mean-variance constrained priors have finite maximum bayes risk in the normal location model. *arXiv preprint arXiv:2303.08653*.
- Chen, J., J. Gu, and S. Kwon (2025). Empirical bayes shrinkage (mostly) does not correct the measurement error in regression. Technical report.
- Chesher, A. (1991). The effect of measurement error. *Biometrika* 78(3), 451–462.
- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review* 104(9), 2593–2632.

- Chetty, R., J. N. Friedman, and J. E. Rockoff (2014b). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review* 104(9), 2633–2679.
- Chetty, R. and N. Hendren (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics* 133(3), 1163–1228.
- Deeb, A. (2021). A framework for using value-added in regressions. *arXiv preprint arXiv:2109.01741*.
- Efron, B. and C. Morris (1973). Stein’s estimation rule and its competitors—an empirical bayes approach. *Journal of the American Statistical Association* 68(341), 117–130.
- Einav, L., A. Finkelstein, and N. Mahoney (2025). Producing health: measuring value added of nursing homes. *Econometrica* 93(4), 1225–1264.
- Evdokimov, K. S. and A. Zeleneev (2019). Errors-in-variables in large nonlinear panel and network models. Technical report.
- Evdokimov, K. S. and A. Zeleneev (2023). Simple estimation of semiparametric models with measurement errors. *arXiv preprint arXiv:2306.14311*.
- Ferguson, T. S. (2017). *A course in large sample theory*. Routledge.
- Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics* 182(1), 156–173.
- Guo, M. and M. Ghosh (2012). Mean squared error of james–stein estimators for measurement error models. *Statistics & Probability Letters* 82(11), 2033–2043.
- Hull, P. (2018). Estimating hospital quality with quasi-experimental data. *Available at SSRN 3118358*.
- Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy* 126(5), 2072–2107.
- Jacob, B. A. and L. Lefgren (2007). What do parents value in education? an empirical investigation of parents’ revealed preferences for teachers. *The Quarterly Journal of Economics* 122(4), 1603–1637.
- Jacob, B. A. and L. Lefgren (2008). Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of labor Economics* 26(1), 101–136.
- Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic

Research.

- Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics* 137(4), 1963–2036.
- Kline, P., R. Saggio, and M. Sølvssten (2020). Leave-out estimation of variance components. *Econometrica* 88(5), 1859–1898.
- Morris, C. N. (1983). Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association* 78(381), 47–55.
- O’Neill, B. (2014). Some useful moment results in sampling problems. *The American Statistician* 68(4), 282–296.
- Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International economic review*, 221–247.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74(4), 967–1012.
- Walters, C. (2024). Empirical bayes methods in labor economics. In *Handbook of Labor Economics*, Volume 5, pp. 183–260. Elsevier.
- Warnick, M., J. Light, and A. Yim (2024). Instructor value-added in higher education.
- Whittemore, A. S. (1989). Errors-in-variables regression using stein estimates. *The American Statistician* 43(4), 226–228.

Supplemental Appendix

B Lemmas and Proofs for Asymptotic Normality

Firstly, for simplicity we abbreviate notations and denote $\text{Var}(\theta_i)$ as V .

Remark B.1. Recall that the variance estimator is defined as

$$\hat{\sigma}_i^2 := \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - \bar{X}_i)^2.$$

Then $\hat{\sigma}_i^2$ is unbiased for σ_i^2 and we have for its variance that

$$\begin{aligned} \text{Var}(\hat{\sigma}_i^2 \mid \sigma_i^2, J_i) &= \frac{1}{J_i} \mathbb{E}[\epsilon_{i,j}^4 \mid \sigma_i^2, J_i] - \frac{\sigma_i^2 (J_i - 3)}{J_i (J_i - 1)} \\ &\leq \frac{1}{J_i} K \sigma_i^4 - \frac{\sigma_i^2 (J_i - 3)}{J_i (J_i - 1)} && \text{(Assumption 3.1.3)} \\ &\leq \frac{1}{J_i} K \sigma_i^4, && \text{(Assumption 3.1.1)} \end{aligned} \quad (6)$$

where the equality follows from e.g. O'Neill (2014) (Result 3, p. 284).

Lemma B.1. *Under Assumption 3.1, we have:*

1. *Properties of $\bar{\epsilon}_i, \bar{X}_i, \hat{\sigma}_i^2, c_i$:*

For any integer $k_1 \in \{1, 2\}$, $k_2 \geq 1$, we have

- (a) $\bar{\epsilon}_i = O_p(n^{-1/4})$, and $\mathbb{E}[|\bar{\epsilon}_i|^{k_1}] \rightarrow 0$.
- (b) $\bar{X}_i = \theta_i + O_p(n^{-1/4})$, and $\mathbb{E}[|\bar{X}_i|^{k_1}] \rightarrow \mathbb{E}[|\theta_i|^{k_1}]$.
- (c) $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$, and $\mathbb{E}[|\hat{\sigma}_i^2 - \sigma_i^2|^{k_1}] \rightarrow 0$.
- (d) $\mathbb{E}\left[\frac{1}{J_i} \hat{\sigma}_i^2\right] \rightarrow 0$, and $\mathbb{E}\left[\left(\frac{1}{J_i} \hat{\sigma}_i^2\right)^2\right] \rightarrow 0$.
- (e) $\mathbb{E}\left[\left(\frac{1}{J_i} \hat{\sigma}_i^2\right)^4\right] = o(n^{-1/2})$.
- (f) $c_i^{k_2} = 1 + O_p(n^{-1/2})$, and $\mathbb{E}[c_i^{k_2}] = 1 + O(n^{-1/2})$.

2. *Properties of sample moments of $\bar{\epsilon}_i, c_i, \theta_i$:*

For any integer $k_1 \geq 0^2$, $0 \leq k_2 \leq 2$, and $k_3, k_4 \in \{0, 1\}$, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} &= \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] + O_p(n^{-1/2}), \\ \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - \mathbb{E}[\theta_i])^{k_4} \bar{\epsilon}_i &= o_p(n^{-1/2}), \\ \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 &= O_p(n^{-1/2}). \end{aligned}$$

3. Properties of sample means of θ_i , \bar{X}_i , $\hat{\sigma}_i^2$:

- (a) $\bar{\theta} = \mathbb{E}[\theta_i] + O_p(n^{-1/2})$.
- (b) $\bar{X} = \mathbb{E}[\theta_i] + O_p(n^{-1/2})$.
- (c) $\frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 = O_p(n^{-1/2})$

4. Properties of sample moments of ϵ_i , c_i , u_i :

For any integer $k \geq 0$, we have

- (a) $\frac{1}{n} \sum_{i=1}^n c_i^k u_i = O_p(n^{-1/2})$.
- (b) $\frac{1}{n} \sum_{i=1}^n c_i^k u_i \bar{\epsilon}_i = o_p(n^{-1/2})$.

Proof of Lemma B.1. Below we take any $s > 0$,

1. Properties of $\bar{\epsilon}_i$, \bar{X}_i , $\hat{\sigma}_i^2$, c_i :

- (a) By Markov's inequality,

$$\Pr(n^{1/4} |\bar{\epsilon}_i| > s) \leq \frac{\sqrt{n} \mathbb{E}[\bar{\epsilon}_i^2]}{s^2} = \frac{\sqrt{n} \mathbb{E}\left[\frac{1}{J_i}\right] \mathbb{E}[\sigma_i^2]}{s^2}.$$

Since $\sqrt{n} \mathbb{E}\left[\frac{1}{J_i}\right] \rightarrow \kappa$, we have $\bar{\epsilon}_i = O_p(n^{-1/4})$.

Since

$$\mathbb{E}[\bar{\epsilon}_i^2] = \mathbb{E}\left[\frac{1}{J_i}\right] \mathbb{E}[\sigma_i^2] \rightarrow 0,$$

²For statements like this, if $k = 0$ is on the exponent, it means that the term is 1.

combined with the Cauchy-Schwarz inequality, we have the rest of the results.

(b) This follows from [1a](#).

(c) By Markov's inequality and [\(6\)](#),

$$\begin{aligned} \Pr \left(n^{1/4} |\hat{\sigma}_i^2 - \sigma_i^2| > s \right) &\leq \frac{\sqrt{n} \mathbb{E} \left[(\hat{\sigma}_i^2 - \sigma_i^2)^2 \right]}{s^2} \\ &\leq \frac{K \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^4]}{s^2}. \end{aligned}$$

Due to Assumption [3.1.7](#), we have $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$.

Since by [\(6\)](#),

$$\mathbb{E} \left[(\hat{\sigma}_i^2 - \sigma_i^2)^2 \right] \leq K \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^4] \rightarrow 0,$$

combined with the Cauchy-Schwarz inequality, we have the rest of the results.

(d) We have

$$\mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \right] = \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^2] \rightarrow 0,$$

$$\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right] \leq \mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K \sigma_i^4}{J_i} + \sigma_i^4 \right) \right] \rightarrow 0. \quad (\text{By } \a href="#">(6))$$

(e) From Theorem 2 of [Angelova \(2012\)](#), the fourth moment of $\hat{\sigma}_i^2$ is

$$\begin{aligned} \mathbb{E} \left[(\hat{\sigma}_i^2)^4 \mid J_i, \sigma_i^2 \right] &= \mu_2^4 + \frac{6\mu_2^2(\mu_4 - \mu_2^2)}{J_i} + \frac{4\mu_6\mu_2 + 3\mu_4^2 - 18\mu_4\mu_2^2 - 24\mu_3^2\mu_2 + 23\mu_2^4}{J_i^2} \\ &\quad + \frac{\mu_8 - 4\mu_6\mu_2 - 24\mu_5\mu_3 - 3\mu_4^2 + 72\mu_4\mu_2^2 + 96\mu_3^2\mu_2 - 86\mu_2^4}{J_i^3} \\ &\quad + \frac{4(6\mu_6\mu_2 + 6\mu_4^2 - 39\mu_4\mu_2^2 - 40\mu_3^2\mu_2 + 45\mu_2^4)}{J_i^3(J_i - 1)} \\ &\quad + \frac{4(36\mu_4\mu_2^2 - 8\mu_5\mu_3 + 52\mu_3^2\mu_2 - 61\mu_2^4)}{J_i^3(J_i - 1)^2} \\ &\quad + \frac{8(\mu_4^2 - 6\mu_4\mu_2^2 - 12\mu_3^2\mu_2 + 15\mu_2^4)}{J_i^3(J_i - 1)^3}, \end{aligned}$$

where $\mu_k, k = 1, \dots, 8$ are the k -th raw moments of $\epsilon_{i,j} \mid \sigma_i^2$. By Assumption [3.1.3](#), we have $\mu_k \leq K\sigma_i^k$ for some constant K . Thus, by independence of Assumption [3.1.1](#) and Assumption [3.1.8](#), we have that

$$\sqrt{n}\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^4 \right] \rightarrow 0.$$

(f) By the mean value theorem,

$$\begin{aligned} c_i^{k_2} &= \left(\frac{V}{V + \frac{1}{J_i} \hat{\sigma}_i^2} \right)^{k_2} \\ &= 1 - k_2 \frac{1}{V J_i} \hat{\sigma}_i^2 + \frac{k_2(k_2 + 1)}{2} \frac{1}{(1 + \omega_i)^{k_2+2}} \frac{1}{V^2 J_i^2} (\hat{\sigma}_i^2)^2, \end{aligned}$$

where ω_i is between 0 and $\frac{1}{\sqrt{J_i}}\hat{\sigma}_i^2$. Therefore,

$$\begin{aligned}
|\sqrt{n}\mathbb{E}[1 - c_i^{k_2}]| &\leq \frac{k_2}{V}\sqrt{n}\mathbb{E}\left[\frac{\sigma_i^2}{J_i}\right] \\
&\quad + \frac{k_2(k_2+1)}{2V^2}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\left(\sigma_i^4 + \frac{K}{J_i}\sigma_i^4\right)\right] && \text{(By (6))} \\
&\leq \frac{k_2}{V}\mathbb{E}[\sigma_i^2]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\right] \\
&\quad + \frac{k_2(k_2+1)}{2V^2}\mathbb{E}[\sigma_i^4]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\right] \\
&\quad + \frac{k_2(k_2+1)K}{2V^2}\mathbb{E}[\sigma_i^4]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^3}\right] && \text{(By Assumption 3.1.1)} \\
&\rightarrow \frac{k_2}{V}\mathbb{E}[\sigma_i^2]\kappa + 0 + 0. && \text{(By Assumption 3.1.7 and 3.1.8)}
\end{aligned}$$

Then we have the second result. By Markov's inequality, we have the first result.

2. By Chebyshev's inequality (k_2 can be replaced by k_4),

$$\begin{aligned}
&\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n c_i^{k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2} - \mathbb{E}\left[c_i^{k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2}\right]\right| > s\right) \\
&\leq \frac{\mathbb{E}\left[c_i^{2k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{2k_2}\right]}{s^2} \leq \frac{\mathbb{E}\left[(\theta_i - k_3\mathbb{E}[\theta_i])^{2k_2}\right]}{s^2} \\
&\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n c_i^{k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2}\bar{\epsilon}_i - \mathbb{E}\left[c_i^{k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2}\bar{\epsilon}_i\right]\right| > s\right) \\
&\leq \frac{\mathbb{E}\left[c_i^{2k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{2k_2}\bar{\epsilon}_i^2\right]}{s^2} \leq \frac{\mathbb{E}\left[(\theta_i - k_3\mathbb{E}[\theta_i])^{2k_2}\right]\mathbb{E}\left[\frac{1}{J_i}\right]\mathbb{E}[\sigma_i^2]}{s^2} \rightarrow 0,
\end{aligned}$$

$$\begin{aligned}
& \Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 - \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 \right] \right| > s \right) \\
& \leq \frac{\mathbb{E} \left[c_i^{2k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \bar{\epsilon}_i^4 \right]}{s^2} \\
& \leq \frac{\mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \right] \left(\mathbb{E} \left[\frac{1}{J_i^3} \right] \mathbb{E} [K \sigma_i^4] + 3 \mathbb{E} \left[\frac{J_i - 1}{J_i^3} \right] \mathbb{E} [\sigma_i^4] \right)}{s^2} \rightarrow 0.
\end{aligned}$$

Meanwhile, by independence from Assumption 3.1.3,

$$\begin{aligned}
& \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] = \mathbb{E} [c_i^{k_1}] \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] \\
& = (1 + O(n^{-1/2})) \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] \quad (\text{By 1f}) \\
& = \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] + O(n^{-1/2}),
\end{aligned}$$

by the mean value theorem ($k_4 = 0$) and (6),

$$\begin{aligned}
|\sqrt{n} \mathbb{E} [c_i^{k_1} \bar{\epsilon}_i]| & \leq 0 + \frac{k_1}{V} \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \bar{\epsilon}_i \right] \\
& \quad + \frac{k_1 (k_1 + 1)}{2V^2} \sqrt{n} \mathbb{E} \left[\frac{1}{J_i^2} (\hat{\sigma}_i^2)^2 \bar{\epsilon}_i \right] \\
& \leq \frac{k_1}{V} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^2} \left(\sigma_i^4 + \frac{K}{J_i} \sigma_i^4 \right) \right]} \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \right] \\
& \quad + \frac{k_1 (k_1 + 1)}{2V^2} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right]} \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \right] \quad (\text{Cauchy-Schwarz}) \\
& \rightarrow 0,
\end{aligned}$$

where the last step follows from independence of Assumption 3.1.1, Assumption 3.1.7, Assumption 3.1.8 and 1e,

by the independence in Assumption 3.1.3 ($k_4 = 1$),

$$\mathbb{E} [c_i^{k_1} (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i] = \mathbb{E} \left[\mathbb{E} \left(c_i^{k_1} (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i \middle| \epsilon \right) \right] = 0,$$

and also by Assumption 3.1.3,

$$\left| \sqrt{n} \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E} [\theta_i])^{k_2} \tilde{\epsilon}_i^2 \right] \right| \leq \mathbb{E} \left[\left| (\theta_i - k_3 \mathbb{E} [\theta_i])^{k_2} \right| \right] \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^2],$$

where $\sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \right] \rightarrow \kappa$. Therefore we have the results.

3. (a) Follows from the finite fourth moment and the inequality below for any $s > 0$,

$$\Pr \left(\left| \sqrt{n} (\bar{\theta} - \mathbb{E} [\theta_i]) \right| > s \right) \leq \frac{\mathbb{E} [\theta_i^2]}{s^2}. \quad (\text{Chebyshev's inequality})$$

- (b) Follows from the property of $\bar{\theta}$ and the second property of 2:

$$\bar{X} = \bar{\theta} + \bar{\epsilon} = \mathbb{E} [\theta_i] + O_p(n^{-1/2}) + o_p(n^{-1/2}) = \mathbb{E} [\theta_i] + O_p(n^{-1/2}).$$

- (c) By Chebyshev's inequality and (6), for any $s > 0$,

$$\Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 - \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \right] \right| > s \right) \leq \frac{\mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K\sigma_i^4}{J_i} + \sigma_i^4 \right) \right]}{s^2} \rightarrow 0,$$

and also

$$\sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \right] = \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^2] \rightarrow \kappa \mathbb{E} [\sigma_i^2].$$

Thus

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 = O_p(n^{-1/2}).$$

4. (a) Since $u_i \perp \epsilon_{i,j} \mid \theta_i$, by Chebyshev's inequality,

$$\Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^k u_i \right| > s \right) \leq \frac{\mathbb{E} [c_i^{2k} u_i^2]}{s^2} \leq \frac{\sigma_u^2}{s^2}$$

Thus

$$\frac{1}{n} \sum_{i=1}^n c_i^k u_i = O_p(n^{-1/2}).$$

(b) By Chebyshev's inequality,

$$\Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^k u_i \bar{\epsilon}_i \right| > s \right) \leq \frac{\mathbb{E} [c_i^{2k} u_i^2 \bar{\epsilon}_i^2]}{s^2} \leq \frac{\sigma_u^2}{s^2} \mathbb{E} \left[\frac{1}{J_i} \right] \mathbb{E} [\sigma_i^2] \rightarrow 0.$$

Thus

$$\frac{1}{n} \sum_{i=1}^n c_i^k u_i \bar{\epsilon}_i = o_p(n^{-1/2}).$$

□

Lemma B.2. *Under Assumption 3.1, we have*

$$\frac{1}{n} \sum_{i=1}^n c_i^2 \bar{\epsilon}_i^2 = \frac{1}{n} \sum_{i=1}^n c_i^2 \frac{1}{J_i} \hat{\sigma}_i^2 + o_p(n^{-1/2}).$$

Proof of Lemma B.2. First, notice that the difference

$$\begin{aligned} \bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 &= \bar{\epsilon}_i^2 - \frac{1}{J_i (J_i - 1)} \sum_{j=1}^{J_i} \epsilon_{i,j}^2 + \frac{1}{J_i - 1} \bar{\epsilon}_i^2 \\ &= \frac{1}{J_i (J_i - 1)} \left(\sum_{j=1}^{J_i} \epsilon_{i,j} \right)^2 - \frac{1}{J_i (J_i - 1)} \sum_{j=1}^{J_i} \epsilon_{i,j}^2 \\ &= \frac{2}{J_i (J_i - 1)} \sum_{k \leq j} \epsilon_{i,k} \epsilon_{i,j}. \end{aligned}$$

Then the properties of the difference are

$$\begin{aligned} \mathbb{E} \left[\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right] &= 0, \\ \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \mid J_i, \sigma_i^2 \right] &= \frac{2\sigma_i^4}{J_i (J_i - 1)}. \end{aligned}$$

For c_i^2 , by the mean value theorem,

$$\begin{aligned} c_i^2 &= \left(\frac{V}{V + \frac{1}{J_i} \hat{\sigma}_i^2} \right)^2 \\ &= 1 - 2 \frac{1}{V J_i} \hat{\sigma}_i^2 + 3 \frac{1}{(1 + \omega_i)^4} \frac{1}{V^2 J_i^2} (\hat{\sigma}_i^2)^2, \end{aligned}$$

where ω_i is between 0 and $\frac{1}{V J_i} \hat{\sigma}_i^2$. Therefore,

$$\begin{aligned} & \left| \sqrt{n} \mathbb{E} \left[c_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\ & \leq \left| \sqrt{n} \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| + \frac{2}{V} \left| \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\ & \quad + \frac{3}{V^2} \left| \sqrt{n} \mathbb{E} \left[\frac{1}{(1 + \omega_i)^4} \frac{1}{J_i^2} (\hat{\sigma}_i^2)^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\ & \leq 0 + \frac{2}{V} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^2} (\hat{\sigma}_i^2)^2 \right] \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]} \\ & \quad + \frac{3}{V^2} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right] \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]} \quad (\text{Cauchy-Schwarz}) \\ & \leq \frac{2}{V} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K}{J_i} \sigma_i^4 + \sigma_i^4 \right) \right]} \sqrt{n} \mathbb{E} \left[\frac{2\sigma_i^4}{J_i (J_i - 1)} \right] \\ & \quad + \frac{3}{V^2} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right]} \sqrt{n} \mathbb{E} \left[\frac{2\sigma_i^4}{J_i (J_i - 1)} \right]. \quad (\text{By properties of the difference}) \end{aligned}$$

Since we have independence of $\sigma_i^2 \perp J_i$ from Assumption 3.1.1, Assumption 3.1.7, Assumption 3.1.8, and that

$$\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right] \rightarrow 0,$$

from Lemma B.1.1e, the above converges to 0. Thus, the expectation

$$\sqrt{n}\mathbb{E}\left[c_i^2\left(\bar{\epsilon}_i^2 - \frac{1}{J_i}\hat{\sigma}_i^2\right)\right] \rightarrow 0.$$

Next, we show the convergence to expectations by Chebyshev's inequality:

$$\begin{aligned} & \Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n c_i^2\left(\bar{\epsilon}_i^2 - \frac{1}{J_i}\hat{\sigma}_i^2\right) - \mathbb{E}\left[c_i^2\left(\bar{\epsilon}_i^2 - \frac{1}{J_i}\hat{\sigma}_i^2\right)\right]\right| > s\right) \\ & \leq \frac{\mathbb{E}\left[c_i^4\left(\bar{\epsilon}_i^2 - \frac{1}{J_i}\hat{\sigma}_i^2\right)^2\right]}{s^2} \leq \frac{\mathbb{E}\left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i}\hat{\sigma}_i^2\right)^2\right]}{s^2} = \frac{\mathbb{E}\left[\frac{2\sigma_i^4}{J_i(J_i-1)}\right]}{s^2} \rightarrow 0. \end{aligned}$$

And the proof is complete. \square

Recall that we define in the beginning of this section $V := \text{Var}(\theta_i)$ and in Section 3.3

$$\hat{V} := \frac{1}{n}\sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{n-1}{n^2}\sum_{i=1}^n \frac{1}{J_i}\hat{\sigma}_i^2.$$

The following lemma shows that \hat{V} converges to V at \sqrt{n} -rate.

Lemma B.3. *Under Assumption 3.1, we have*

$$\hat{V} = V + O_p(n^{-1/2}).$$

Proof of Lemma B.3. By Lemma B.1,

$$\begin{aligned} \hat{V} &= \frac{1}{n}\sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{n-1}{n^2}\sum_{i=1}^n \frac{1}{J_i}\hat{\sigma}_i^2 \\ &= \frac{1}{n}\sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{n}\sum_{i=1}^n \bar{\epsilon}_i^2 - (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \bar{\epsilon}^2 + \frac{2}{n}\sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])\bar{\epsilon}_i \\ &\quad - 2(\bar{\theta} - \mathbb{E}[\theta_i])\bar{\epsilon} - \frac{n-1}{n^2}\sum_{i=1}^n \frac{1}{J_i}\hat{\sigma}_i^2 \\ &= V + O_p(n^{-1/2}) + O_p(n^{-1/2}) - O_p(n^{-1}) - o_p(n^{-1}) + o_p(n^{-1/2}) \\ &\quad - o_p(n^{-1}) - O_p(n^{-1/2}) \\ &= V + O_p(n^{-1/2}). \end{aligned}$$

□

We prove the common-weight result here.

Proof of Proposition 3.5. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{\text{CW}}$ as $\hat{\beta}$, and $\text{Var}(\theta_i)$ as V .

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &= \hat{V}^{-1} \left(\beta \frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{X}_i - \bar{X}) (\theta_i - \bar{\theta}) - \beta \sqrt{n} \hat{V} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\bar{X}_i - \bar{X}) (u_i - \bar{u}) \right) \\ &= \frac{\beta \sqrt{n} T_{1,n} - \beta \sqrt{n} T_{2,n} + \sqrt{n} T_{3,n}}{T_{2,n}}.\end{aligned}$$

Firstly, from Lemma B.1 and from the proof of Lemma B.2, we have for the denominator,

$$\begin{aligned}T_{2,n} &= \frac{1}{n} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 + o_p(n^{-1/2}) \\ &= \frac{1}{n} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{1}{n} \sum_{i=1}^n \bar{\epsilon}_i^2 + o_p(n^{-1/2}).\end{aligned}$$

From Lemma B.3,

$$T_{2,n} = V + O_p(n^{-1/2}).$$

For the numerator terms, by properties from Lemma B.1,

$$\begin{aligned}\sqrt{n} T_{1,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i (\theta_i - \mathbb{E}[\theta_i]) \\ &\quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{\epsilon} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + o_p(1),\end{aligned}$$

$$\begin{aligned}
\sqrt{n}T_{3,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i u_i \\
&\quad - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{u} - \sqrt{n} \bar{\epsilon} \bar{u} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + o_p(1).
\end{aligned}$$

Combined with the proof of Lemma B.3, we have

$$\begin{aligned}
\sqrt{n}T_{2,n} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [(\theta_i - \mathbb{E}[\theta_i])^2 + \bar{\epsilon}_i^2] - \sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \sqrt{n} \bar{\epsilon}^2 + \frac{2}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i \\
&\quad - 2\sqrt{n} (\bar{\theta} - \mathbb{E}[\theta_i]) \bar{\epsilon} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \bar{\epsilon}_i^2 \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + o_p(1).
\end{aligned}$$

Therefore, the numerator is

$$\begin{aligned}
&\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) u_i + o_p(1).
\end{aligned}$$

Then applying the central limit theorem combined with the denominator, by Slutsky's theorem,

$$\sqrt{n} (\hat{\beta} - \beta) = \frac{\beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n}}{T_{2,n}} \rightarrow_d N \left(0, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2]}{V^2} \right).$$

□

C Lemmas and Proofs for Inference

Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{c,HE}$ as $\hat{\beta}_c$, $\hat{\beta}_{HE}$ as $\hat{\beta}$, $\hat{\theta}_{i,c,HE}$ as $\hat{\theta}_{i,c}$, $\hat{\theta}_{i,HE}$ as $\hat{\theta}_i$ and $\text{Var}(\theta_i)$ as V .

Lemma C.1. *Under Assumption 3.1 and Assumption 3.2,*

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 \hat{u}_{i,c}^2 \rightarrow_p \mathbb{E} \left[u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2 \right].$$

Proof of Lemma C.1. Combining Lemma C.2, C.3, C.4, C.5, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 \hat{u}_{i,c}^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 \left(Y_i - \bar{Y} - \hat{\beta}_c \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (Y_i - \bar{Y})^2 - 2\hat{\beta}_c \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^3 (Y_i - \bar{Y}) \\ & \quad + \hat{\beta}_c^2 \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^4. \\ &= \beta^2 \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (\theta_i - \bar{\theta})^2 - 2\beta \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (\theta_i - \bar{\theta}) (u_i - \bar{u}) \\ & \quad + \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (u_i - \bar{u})^2 \\ & \quad - 2\hat{\beta}_c \beta \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^3 (\theta_i - \bar{\theta}) + 2\hat{\beta}_c \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^3 (u_i - \bar{u}) \\ & \quad + \hat{\beta}_c^2 \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^4 \\ &= \beta^2 \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^4 \right] - 2\beta \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^3 u_i \right] + \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^2 u_i^2 \right] \\ & \quad - 2\beta^2 \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^4 \right] + 2\beta \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^3 u_i \right] + \beta^2 \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^4 \right] + o_p(1) \\ &= \mathbb{E} \left[u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2 \right] + o_p(1). \end{aligned}$$

□

Lemma C.2. *Under Assumption 3.1 and Assumption 3.2,*

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^4 \rightarrow_p \mathbb{E} \left[(\theta_i - \mathbb{E}[\theta_i])^4 \right].$$

Proof of Lemma C.2. For any $s > 0$ and integers $k_1 \geq 0$, $0 \leq k_2, k_3 \leq 4$,

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_2} - \mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2}] \right| > s \right) \leq \frac{\mathbb{E} [c_i^{2k_1} \bar{X}_i^{2k_2}]}{ns^2} \leq \frac{\mathbb{E} [\bar{X}_i^{2k_2}]}{ns^2} \rightarrow 0.$$

$$\mathbb{E} [c_i^{k_1} \theta_i^{k_3}] = \mathbb{E} [c_i^{k_1}] \mathbb{E} [\theta_i^{k_3}] \rightarrow \mathbb{E} [\theta_i^{k_3}]. \quad (\text{By Lemma B.1 1f})$$

For $k_4 = 4$, by independence and (6),

$$\begin{aligned} |\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \bar{\epsilon}_i^{k_4}]| &\leq \mathbb{E} [|\theta_i^{k_3}|] \mathbb{E} [\bar{\epsilon}_i^{k_4}] = \mathbb{E} [|\theta_i^{k_3}|] \mathbb{E} \left[\frac{1}{J_i^3} \mathbb{E} [\epsilon_{i,j}^4 | J_i, \sigma_i^2] + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\ &\leq \mathbb{E} [|\theta_i^{k_3}|] \mathbb{E} \left[\frac{1}{J_i^3} K \sigma_i^4 + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\ &= \mathbb{E} [|\theta_i^{k_3}|] \left(\mathbb{E} \left[\frac{1}{J_i^3} \right] K \mathbb{E} [\sigma_i^4] + \mathbb{E} \left[\frac{3(J_i - 1)}{J_i^3} \right] \mathbb{E} [\sigma_i^2] \right) \rightarrow 0. \end{aligned}$$

Then by Jensen's inequality, for $1 \leq k_4 \leq 4$,

$$|\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \bar{\epsilon}_i^{k_4}]| \rightarrow 0.$$

Therefore for $0 \leq k_2 \leq 4$,

$$\begin{aligned} &\mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2}] \\ &= \sum_{k_4=0}^{k_2} \binom{k_2}{k_4} \mathbb{E} [c_i^{k_1} \theta_i^{k_2-k_4} \bar{\epsilon}_i^{k_4}] \\ &\rightarrow \mathbb{E} [\theta_i^{k_2}]. \end{aligned}$$

So for $0 \leq k_2 \leq 4$,

$$\frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_2} \rightarrow_p \mathbb{E} [\theta_i^{k_2}].$$

Therefore for $1 \leq k \leq 4$, $k_2 \geq 0$,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^k \\
&= \sum_{k_1=0}^k \binom{k}{k_1} \left(-\frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^{k-k_1} \frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_1} \\
&= \sum_{k_1=0}^k \binom{k}{k_1} (-\mathbb{E}[\theta_i])^{k-k_1} \mathbb{E}[\theta_i^{k_1}] + o_p(1) \\
&= \mathbb{E}[(\theta_i - \mathbb{E}[\theta_i])^k] + o_p(1).
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^k c_i^{k_2} \\
&= \sum_{k_1=0}^k \binom{k}{k_1} \left(-\frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^{k-k_1} \frac{1}{n} \sum_{i=1}^n c_i^{k_1+k_2} \bar{X}_i^{k_1} \\
&= \mathbb{E}[(\theta_i - \mathbb{E}[\theta_i])^k] + o_p(1).
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^k (c_i - \bar{c})^{k_2} \\
&= \sum_{k_1=0}^k \binom{k}{k_1} \left(-\frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^{k-k_1} \frac{1}{n} \sum_{i=1}^n c_i^{k_1+k_2} \bar{X}_i^{k_1} (c_i - \bar{c})^{k_2} \\
&= \sum_{k_1=0}^k \binom{k}{k_1} \left(-\frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^{k-k_1} \sum_{t=0}^{k_2} \binom{k_2}{t} (-\bar{c})^{k_2-t} \frac{1}{n} \sum_{i=1}^n c_i^{k_1+k_2+t} \bar{X}_i^{k_1} \\
&= \sum_{k_1=0}^k \binom{k}{k_1} (-\mathbb{E}[\theta_i])^{k-k_1} \sum_{t=0}^{k_2} \binom{k_2}{t} (-1)^{k_2-t} \mathbb{E}[\theta_i^{k_1}] + o_p(1) \\
&= o_p(1).
\end{aligned}$$

The above also applies to $k = 0$.

Therefore,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^4 \\
&= \frac{1}{n} \sum_{i=1}^n \left(c_i (\bar{X}_i - \bar{X}) - \frac{1}{n} \sum_{i=1}^n (c_i (\bar{X}_i - \bar{X})) \right)^4 \\
&= \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i - \bar{X} (c_i - \bar{c}) \right)^4 \\
&= \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^4 - 4\bar{X} \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^3 (c_i - \bar{c}) \\
&\quad + 6\bar{X}^2 \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right)^2 (c_i - \bar{c})^2 - 4\bar{X}^3 \frac{1}{n} \sum_{i=1}^n \left(c_i \bar{X}_i - \frac{1}{n} \sum_{i=1}^n c_i \bar{X}_i \right) (c_i - \bar{c})^3 \\
&\quad + \bar{X}^4 \frac{1}{n} \sum_{i=1}^n (c_i - \bar{c})^4. \\
&= \mathbb{E} [(\theta_i - \mathbb{E}[\theta_i])^4] + o_p(1).
\end{aligned}$$

□

Lemma C.3. *Under Assumption 3.1 and Assumption 3.2,*

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^k (\theta_i - \bar{\theta})^{4-k} \rightarrow_p \mathbb{E} [(\theta_i - \mathbb{E}[\theta_i])^4].$$

Proof of Lemma C.3. For any $s > 0$ and integers $k_1 \geq 0$, $0 \leq k_2 \leq 4$, $k_2 + k_3 \leq 4$, by Chebyshev's inequality,

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_2} \theta_i^{k_3} - \mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} \theta_i^{k_3}] \right| > s \right) \leq \frac{\mathbb{E} [c_i^{2k_1} \bar{X}_i^{2k_2} \theta_i^{2k_3}]}{ns^2} \leq \frac{\mathbb{E} [\bar{X}_i^{2k_2} \theta_i^{2k_3}]}{ns^2} \rightarrow 0.$$

For $0 \leq k_2 \leq 4$, $k_2 + k_3 \leq 4$,

$$\begin{aligned}
& \mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} \theta_i^{k_3}] \\
&= \sum_{k_4=0}^{k_2} \binom{k_2}{k_4} \mathbb{E} [c_i^{k_1} \theta_i^{k_2-k_4+k_3} \bar{\epsilon}_i^{k_4}] \\
&\rightarrow \mathbb{E} [\theta_i^{k_2+k_3}]. \tag{By Lemma B.1 1f}
\end{aligned}$$

Therefore, similar to the proof of Lemma C.2,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,c} - \bar{\theta}_c)^k (\theta_i - \bar{\theta})^{4-k} \\
&= \sum_{k_1=0}^{4-k} \binom{4-k}{k_1} \bar{\theta}^{4-k-k_1} \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,c} - \bar{\theta}_c)^k \theta_i^{k_1} \\
&= \sum_{k_1=0}^{4-k} \binom{4-k}{k_1} (\mathbb{E} [\theta_i])^{4-k-k_1} \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^k \theta_i^{k_1}] + o_p(1) \\
&= \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^4] + o_p(1).
\end{aligned}$$

□

Lemma C.4. *Under Assumption 3.1 and Assumption 3.2,*

$$\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,c} - \bar{\theta}_c)^3 (u_i - \bar{u}) \rightarrow_p \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^3 u_i],$$

$$\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_{i,c} - \bar{\theta}_c)^2 (\theta_i - \bar{\theta}) (u_i - \bar{u}) \rightarrow_p \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^3 u_i].$$

Proof. For any $s > 0$ and integers $k_1 \geq 0$, $0 \leq k_2 \leq 3$, by Chebyshev's inequality,

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_2} u_i - \mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} u_i] \right| > s \right) \leq \frac{\mathbb{E} [\bar{X}_i^{2k_2} u_i^2]}{ns^2} \rightarrow 0. \tag{7}$$

For $0 \leq k_3 \leq 3$,

$$\mathbb{E} [c_i^{k_1} \theta_i^{k_3} u_i] = \mathbb{E} [c_i^{k_1}] \mathbb{E} [\theta_i^{k_3} u_i] \rightarrow \mathbb{E} [\theta_i^{k_3} u_i]. \tag{By Lemma B.1 1f}$$

For $k_4 = 4$, by independence and (6),

$$\begin{aligned}
|\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \bar{\epsilon}_i^{k_4} u_i^2]| &\leq \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} [\bar{\epsilon}_i^{k_4}] = \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} \left[\frac{1}{J_i^3} \mathbb{E} [\epsilon_{i,j}^4 \mid J_i, \sigma_i^2] + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\
&\leq \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} \left[\frac{1}{J_i^3} K \sigma_i^4 + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\
&= \mathbb{E} [|\theta_i^{k_3} u_i^2|] \left(\mathbb{E} \left[\frac{1}{J_i^3} \right] K \mathbb{E} [\sigma_i^4] + \mathbb{E} \left[\frac{3(J_i - 1)}{J_i^3} \right] \mathbb{E} [\sigma_i^2] \right) \rightarrow 0.
\end{aligned}$$

Then by Jensen's inequality, for $1 \leq k_4 \leq 4$,

$$|\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \bar{\epsilon}_i^{k_4} u_i^2]| \rightarrow 0.$$

Therefore for $1 \leq k_2 \leq 4$,

$$\begin{aligned}
&\mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} u_i] \\
&= \sum_{k_4=0}^{k_2} \binom{k_2}{k_4} \mathbb{E} [c_i^{k_1} \theta_i^{k_2-k_4} \bar{\epsilon}_i^{k_4} u_i] \\
&\rightarrow \mathbb{E} [\theta_i^{k_2} u_i].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^3 (u_i - \bar{u}) \\
&= \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^3 u_i + o_p(1) \\
&= \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^3 u_i] + o_p(1).
\end{aligned}$$

The second result is shown by combining the proof procedures of Lemma C.3 and the above. \square

Lemma C.5. *Under Assumption 3.1 and Assumption 3.2,*

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (u_i - \bar{u})^2 \rightarrow_p \mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]$$

Proof. For any $s > 0$ and integers $k_1 \geq 0$, $0 \leq k_2 \leq 3$, by Chebyshev's inequality,

$$\Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} \bar{X}_i^{k_2} u_i^2 - \mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} u_i^2] \right| > s \right) \leq \frac{\mathbb{E} [\bar{X}_i^{2k_2} u_i^4]}{ns^2} \rightarrow 0.$$

For $0 \leq k_3 \leq 3$,

$$\mathbb{E} [c_i^{k_1} \theta_i^{k_3} u_i^2] = \mathbb{E} [c_i^{k_1}] \mathbb{E} [\theta_i^{k_3} u_i^2] \rightarrow \mathbb{E} [\theta_i^{k_3} u_i^2]. \quad (\text{By Lemma B.1 1f})$$

For $k_4 = 4$, by independence and (6),

$$\begin{aligned} |\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \epsilon_i^{k_4} u_i^2]| &\leq \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} [\epsilon_i^{k_4}] = \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} \left[\frac{1}{J_i^3} \mathbb{E} [\epsilon_{i,j}^4 | J_i, \sigma_i^2] + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\ &\leq \mathbb{E} [|\theta_i^{k_3} u_i^2|] \mathbb{E} \left[\frac{1}{J_i^3} K \sigma_i^4 + \frac{3(J_i - 1)}{J_i^3} \sigma_i^2 \right] \\ &= \mathbb{E} [|\theta_i^{k_3} u_i^2|] \left(\mathbb{E} \left[\frac{1}{J_i^3} \right] K \mathbb{E} [\sigma_i^4] + \mathbb{E} \left[\frac{3(J_i - 1)}{J_i^3} \right] \mathbb{E} [\sigma_i^2] \right) \rightarrow 0. \end{aligned}$$

Then by Jensen's inequality, for $1 \leq k_4 \leq 4$,

$$|\mathbb{E} [c_i^{k_1} \theta_i^{k_3} \epsilon_i^{k_4} u_i^2]| \rightarrow 0.$$

Therefore for $1 \leq k_2 \leq 4$,

$$\begin{aligned} &\mathbb{E} [c_i^{k_1} \bar{X}_i^{k_2} u_i^2] \\ &= \sum_{k_4=0}^{k_2} \binom{k_2}{k_4} \mathbb{E} [c_i^{k_1} \theta_i^{k_2-k_4} \epsilon_i^{k_4} u_i^2] \\ &\rightarrow \mathbb{E} [\theta_i^{k_2} u_i^2]. \end{aligned}$$

Therefore,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 (u_i - \bar{u})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \left(\hat{\theta}_{i,c} - \bar{\theta}_c \right)^2 u_i^2 + o_p(1) \\ &= \mathbb{E} [(\theta_i - \mathbb{E} [\theta_i])^2 u_i^2] + o_p(1). \end{aligned}$$

□

D Lemmas and Proofs for Correlated J_i and σ_i^2

Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{c,\text{HE}}$ as $\hat{\beta}_c$, $\hat{\beta}_{\text{HE}}$ as $\hat{\beta}$, $\hat{\theta}_{i,c,\text{HE}}$ as $\hat{\theta}_{i,c}$, $\hat{\theta}_{i,\text{HE}}$ as $\hat{\theta}_i$ and $\text{Var}(\theta_i)$ as V . We keep the subscripts for those related to HO.

We next define the shrinkage weight

$$c_{i,\text{HO}} := \frac{V}{\frac{\mathbb{E}[\sigma_i^2]}{J_i} + V},$$

where $V = \text{Var}(\theta_i)$. And without affecting the results, we define

$$\hat{\theta}_{i,\text{HO}} := \frac{\hat{\sigma}_\theta^2}{\frac{1}{J_i}\hat{\sigma}^2 + \hat{\sigma}_\theta^2}\bar{X}_i + \frac{\frac{1}{J_i}\hat{\sigma}^2}{\frac{1}{J_i}\hat{\sigma}^2 + \hat{\sigma}_\theta^2}\bar{X}.$$

Assumption D.1. 1. J_i is independent of θ_i and u_i . $J_i \perp \epsilon_{i,j} \mid \sigma_i^2$. $J_i \geq 3$, a.s.

2. $\mathbb{E}[u_i] = 0$, $\mathbb{E}(u_i\theta_i) = 0$. $\mathbb{E}[Y_i^4] < \infty$.

3. $\mathbb{E}[\epsilon_{i,j} \mid \sigma_i^2] = 0$, $\mathbb{E}[\epsilon_{i,j}^2 \mid \sigma_i^2] = \sigma_i^2$, $\mathbb{E}[|\epsilon_{i,j}|^L \mid \sigma_i^2] \leq K\sigma_i^L$, $L \geq 3$. Also, $\epsilon_{i,j} \perp \theta_i$.

4. $u_i \perp \epsilon_{i,j} \mid \theta_i$.

5. $\mathbb{E}[\theta_i^4] < \infty$.

6. $\mathbb{E}[\sigma_i^{32}] < \infty$.

Assumption D.2.

$$n^{\frac{3}{2}}\mathbb{E}\left[\frac{1}{J_i^3}\right] \rightarrow \kappa^3$$

Remark D.1. Since we assume $J_i \geq 3$, Assumption D.2 also implies that

$$n\mathbb{E}\left[\frac{1}{J_i^2}\right] \leq \left(n^{\frac{3}{2}}\mathbb{E}\left[\frac{1}{J_i^3}\right]\right)^{2/3} = O(1). \quad (\text{Jensen's inequality})$$

Lemma D.1. Under Assumption D.1, Assumption D.2, we have:

1. Properties of $\bar{\epsilon}_i, \bar{X}_i, \hat{\sigma}_i^2, c_i, c_{i,HO}$:

For any integer $k_1 \in \{1, 2\}$, $k_2 \geq 1$, we have

- (a) $\bar{\epsilon}_i = O_p(n^{-1/4})$, and $\mathbb{E} \left[|\bar{\epsilon}_i|^{k_1} \right] \rightarrow 0$.
- (b) $\bar{X}_i = \theta_i + O_p(n^{-1/4})$, and $\mathbb{E} \left[|\bar{X}_i|^{k_1} \right] \rightarrow \mathbb{E} \left[|\theta_i|^{k_1} \right]$.
- (c) $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$, and $\mathbb{E} \left[|\hat{\sigma}_i^2 - \sigma_i^2|^{k_1} \right] \rightarrow 0$.
- (d) $\mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \right] \rightarrow 0$, and $\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right] \rightarrow 0$.
- (e) $\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^4 \right] = o(n^{-1/2})$.
- (f) $c_i^{k_2} = 1 + O_p(n^{-1/2})$, and $\mathbb{E} \left[c_i^{k_2} \right] = 1 + O(n^{-1/2})$.
- (g) $c_{i,HO}^{k_2} = 1 + O_p(n^{-1/2})$, and $\mathbb{E} \left[c_{i,HO}^{k_2} \right] = 1 + O(n^{-1/2})$.

2. Properties of sample moments of $\bar{\epsilon}_i, c_i, \theta_i, c_{i,HO}$:

For any integer $k_1 \geq 0^3$, $0 \leq k_2 \leq 2$, and $k_3, k_4 \in \{0, 1\}$, we have

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} &= \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] + O_p(n^{-1/2}), \\ \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - \mathbb{E}[\theta_i])^{k_4} \bar{\epsilon}_i &= o_p(n^{-1/2}), \\ \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 &= O_p(n^{-1/2}). \end{aligned}$$

And similar results hold for $c_{i,HO}$.

3. Properties of sample means of $\theta_i, \bar{X}_i, \hat{\sigma}_i^2$:

- (a) $\bar{\theta} = \mathbb{E}[\theta_i] + O_p(n^{-1/2})$.
- (b) $\bar{X} = \mathbb{E}[\theta_i] + O_p(n^{-1/2})$.
- (c) $\frac{1}{n} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 = O_p(n^{-1/2})$

4. Properties of sample moments of $\epsilon_i, c_i, u_i, c_{i,HO}$:

For any integer $k \geq 0$, we have

³For statements like this, if $k = 0$ is on the exponent, it means that the term is 1.

$$(a) \frac{1}{n} \sum_{i=1}^n c_i^k u_i = O_p(n^{-1/2}).$$

$$(b) \frac{1}{n} \sum_{i=1}^n c_i^k u_i \bar{\epsilon}_i = o_p(n^{-1/2}).$$

And similar results hold for $c_{i,\text{HO}}$.

Proof of Lemma D.1. Below we take any $s > 0$,

1. Properties of $\bar{\epsilon}_i, \bar{X}_i, \hat{\sigma}_i^2, c_i$:

(a) By Markov's inequality,

$$\Pr(n^{1/4} |\bar{\epsilon}_i| > s) \leq \frac{\sqrt{n} \mathbb{E}[\bar{\epsilon}_i^2]}{s^2} = \frac{\sqrt{n} \mathbb{E}\left[\frac{1}{J_i} \sigma_i^2\right]}{s^2} \leq \frac{\sqrt{n \mathbb{E}\left[\frac{1}{J_i^2}\right] \mathbb{E}[\sigma_i^4]}}{s^2}.$$

Since $n \mathbb{E}\left[\frac{1}{J_i^2}\right] \rightarrow \kappa^2$, we have $\bar{\epsilon}_i = O_p(n^{-1/4})$.

Since

$$\mathbb{E}[\bar{\epsilon}_i^2] = \mathbb{E}\left[\frac{1}{J_i} \sigma_i^2\right] \leq \sqrt{\mathbb{E}\left[\frac{1}{J_i^2}\right] \mathbb{E}[\sigma_i^4]} \rightarrow 0,$$

combined with the Cauchy-Schwarz inequality, we have the rest of the results.

(b) This follows from [1a](#).

(c) By Markov's inequality and [\(6\)](#),

$$\Pr(n^{1/4} |\hat{\sigma}_i^2 - \sigma_i^2| > s) \leq \frac{\sqrt{n} \mathbb{E}\left[(\hat{\sigma}_i^2 - \sigma_i^2)^2\right]}{s^2} \leq \frac{2\sqrt{n} \mathbb{E}\left[\frac{K}{J_i} \sigma_i^4\right]}{s^2} \leq \frac{2K \sqrt{n \mathbb{E}\left[\frac{1}{(J_i)^2}\right] \mathbb{E}[\sigma_i^8]}}{s^2}.$$

Due to Assumption [D.2](#), we have $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$.

Also,

$$\mathbb{E}\left[(\hat{\sigma}_i^2 - \sigma_i^2)^2\right] \leq 2\mathbb{E}\left[\frac{K}{J_i} \sigma_i^4\right] \leq 2K \sqrt{\mathbb{E}\left[\frac{1}{(J_i)^2}\right] \mathbb{E}[\sigma_i^8]} \rightarrow 0,$$

combined with the Cauchy-Schwarz inequality, we have the rest of the results.

(d) We have

$$0 \leq \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \right] = \mathbb{E} \left[\frac{1}{J_i} \sigma_i^2 \right] \leq \sqrt{\mathbb{E} \left[\frac{1}{J_i^2} \right] \mathbb{E} [\sigma_i^4]} \rightarrow 0,$$

$$\mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right] = \mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K \sigma_i^4}{J_i} + \sigma_i^4 \right) \right] \rightarrow 0.$$

(e) From Theorem 2 of [Angelova \(2012\)](#), the fourth moment of $\hat{\sigma}_i^2$ is

$$\begin{aligned} \mathbb{E} \left[(\hat{\sigma}_i^2)^4 \mid J_i, \sigma_i^2 \right] &= \mu_2^4 + \frac{6\mu_2^2(\mu_4 - \mu_2^2)}{J_i} + \frac{4\mu_6\mu_2 + 3\mu_4^2 - 18\mu_4\mu_2^2 - 24\mu_3^2\mu_2 + 23\mu_2^4}{J_i^2} \\ &\quad + \frac{\mu_8 - 4\mu_6\mu_2 - 24\mu_5\mu_3 - 3\mu_4^2 + 72\mu_4\mu_2^2 + 96\mu_3^2\mu_2 - 86\mu_2^4}{J_i^3} \\ &\quad + \frac{4(6\mu_6\mu_2 + 6\mu_4^2 - 39\mu_4\mu_2^2 - 40\mu_3^2\mu_2 + 45\mu_2^4)}{J_i^3(J_i - 1)} \\ &\quad + \frac{4(36\mu_4\mu_2^2 - 8\mu_5\mu_3 + 52\mu_3^2\mu_2 - 61\mu_2^4)}{J_i^3(J_i - 1)^2} \\ &\quad + \frac{8(\mu_4^2 - 6\mu_4\mu_2^2 - 12\mu_3^2\mu_2 + 15\mu_2^4)}{J_i^3(J_i - 1)^3}, \end{aligned}$$

where $\mu_k, k = 1, \dots, 8$ are the k -th raw moments of $\epsilon_{i,j} \mid \sigma_i^2$. By Assumption [D.1.3](#), we have $\mu_k \leq K\sigma_i^k$ for some constant K . By the law of iterated expectations and Assumption [D.2](#), for the first term we have

$$\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} \sigma_i^8 \right] \leq \sqrt{n \mathbb{E} \left[\frac{1}{J_i^8} \right] \mathbb{E} [\sigma_i^{16}]} \leq \sqrt{n \mathbb{E} \left[\frac{1}{J_i^3} \right] \mathbb{E} [\sigma_i^{16}]} \rightarrow 0.$$

For the rest of the terms, we have the same result. Therefore,

$$\sqrt{n} \mathbb{E} \left[\left(\frac{1}{J_i} \hat{\sigma}_i^2 \right)^4 \right] = o(n^{-1/2}).$$

(f) By the mean value theorem,

$$\begin{aligned} c_i^{k_2} &= \left(\frac{V}{V + \frac{1}{J_i} \hat{\sigma}_i^2} \right)^{k_2} \\ &= 1 - k_2 \frac{1}{V J_i} \hat{\sigma}_i^2 + \frac{k_2 (k_2 + 1)}{2} \frac{1}{(1 + \omega_i)^{k_2+2}} \frac{1}{V^2 J_i^2} (\hat{\sigma}_i^2)^2, \end{aligned}$$

where ω_i is between 0 and $\frac{1}{V J_i} \hat{\sigma}_i^2$. Therefore,

$$\begin{aligned} |\sqrt{n} \mathbb{E} [1 - c_i^{k_2}]| &\leq \frac{k_2}{V} \sqrt{n} \mathbb{E} \left[\frac{\sigma_i^2}{J_i} \right] \\ &\quad + \frac{k_2 (k_2 + 1)}{2V^2} \sqrt{n} \mathbb{E} \left[\frac{1}{J_i^2} \left(\sigma_i^4 + \frac{K}{J_i} \sigma_i^4 \right) \right] \quad (\text{By (6)}) \\ &\leq \frac{k_2}{V} \sqrt{\mathbb{E} [\sigma_i^4] n \mathbb{E} \left[\frac{1}{J_i^2} \right]} \\ &\quad + \frac{k_2 (k_2 + 1)}{2V^2} \sqrt{\mathbb{E} [\sigma_i^8] n \mathbb{E} \left[\frac{1}{J_i^4} \right]} \\ &\quad + \frac{k_2 (k_2 + 1) K}{2V^2} \sqrt{\mathbb{E} [\sigma_i^8] n \mathbb{E} \left[\frac{1}{J_i^6} \right]} \quad (\text{Cauchy-Schwarz}) \\ &= O(1) + o(1) + o(1) \quad (\text{By Assumption D.2}) \end{aligned}$$

Then we have the second result. By Markov's inequality, we have the first result.

(g) The proof is similar to 1f. By the mean value theorem,

$$\begin{aligned} c_{i,\text{HO}}^{k_2} &= \left(\frac{V}{V + \frac{\mathbb{E}[\sigma_i^2]}{J_i}} \right)^{k_2} \\ &= 1 - k_2 \frac{\mathbb{E}[\sigma_i^2]}{V J_i} + \frac{k_2 (k_2 + 1)}{2} \frac{1}{(1 + \omega_i)^{k_2+2}} \frac{\mathbb{E}[\sigma_i^2]^2}{V^2 J_i^2}. \end{aligned}$$

It is easy to see that the rest of the proof is similar to 1f.

2. By Chebyshev's inequality (k_2 can be replaced by k_4),

$$\begin{aligned}
& \Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} - \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] \right| > s \right) \\
& \leq \frac{\mathbb{E} \left[c_i^{2k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \right]}{s^2} \leq \frac{\mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \right]}{s^2} \\
& \Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i - \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i \right] \right| > s \right) \\
& \leq \frac{\mathbb{E} \left[c_i^{2k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \bar{\epsilon}_i^2 \right]}{s^2} \leq \frac{\mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \right] \sqrt{\mathbb{E} \left[\frac{1}{J_i^2} \right] \mathbb{E}[\sigma_i^4]}}{s^2} \rightarrow 0, \\
& \Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 - \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \bar{\epsilon}_i^2 \right] \right| > s \right) \\
& \leq \frac{\mathbb{E} \left[c_i^{2k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \bar{\epsilon}_i^4 \right]}{s^2} \\
& \leq \frac{\mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{2k_2} \right] \left(\sqrt{\mathbb{E} \left[\frac{1}{J_i^6} \right] \mathbb{E} [K^2 \sigma_i^8]} + \sqrt{\mathbb{E} \left[\frac{(J_i - 1)^2}{J_i^6} \right] \mathbb{E} [\sigma_i^8]} \right)}{s^2} \rightarrow 0.
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
& \sqrt{n} \mathbb{E} \left[c_i^{k_1} (\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] = \sqrt{n} \mathbb{E} \left[c_i^{k_1} \right] \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] \\
& = \sqrt{n} \mathbb{E} \left[(\theta_i - k_3 \mathbb{E}[\theta_i])^{k_2} \right] + O(1), \tag{By 1f}
\end{aligned}$$

by the mean value theorem ($k_4 = 0$) and (6),

$$\begin{aligned}
|\sqrt{n}\mathbb{E}[c_i^{k_1}\bar{\epsilon}_i]| &\leq 0 + \frac{k_1}{V}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2\bar{\epsilon}_i\right] \\
&\quad + \frac{k_1(k_1+1)}{2V^2}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}(\hat{\sigma}_i^2)^2\bar{\epsilon}_i\right] \\
&\leq \frac{k_1}{V}\sqrt{\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\left(\sigma_i^4 + \frac{K}{J_i}\sigma_i^4\right)\right]}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\sigma_i^2\right] \\
&\quad + \frac{k_1(k_1+1)}{2V^2}\sqrt{\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^4}(\hat{\sigma}_i^2)^4\right]}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\sigma_i^2\right] \tag{Cauchy-Schwarz} \\
&\leq \frac{k_1}{V}\sqrt{\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^4}\right]}\mathbb{E}[\sigma_i^8] + n\mathbb{E}\left[\frac{1}{J_i^6}\right]K^2\mathbb{E}[\sigma_i^8]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\right]}\mathbb{E}[\sigma_i^4] \\
&\quad + \frac{k_1(k_1+1)}{2V^2}\sqrt{\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^4}(\hat{\sigma}_i^2)^4\right]}\sqrt{n\mathbb{E}\left[\frac{1}{J_i^2}\right]}\mathbb{E}[\sigma_i^4] \\
&= o(1) \cdot O(1) + o(1) \cdot O(1)
\end{aligned}$$

where the last step follows from Assumption D.2 and 1e,

by the independence in Assumption 3.1.3 ($k_4 = 1$),

$$\mathbb{E}[c_i^{k_1}(\theta_i - \mathbb{E}[\theta_i])\bar{\epsilon}_i] = \mathbb{E}\left[\mathbb{E}\left(c_i^{k_1}(\theta_i - \mathbb{E}[\theta_i])\bar{\epsilon}_i \middle| \epsilon\right)\right] = 0,$$

and also by Assumption 3.1.3,

$$\left|\sqrt{n}\mathbb{E}\left[c_i^{k_1}(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2}\bar{\epsilon}_i^2\right]\right| \leq \mathbb{E}\left[|(\theta_i - k_3\mathbb{E}[\theta_i])^{k_2}|\right]\sqrt{n\mathbb{E}\left[\frac{1}{J_i^2}\right]}\mathbb{E}[\sigma_i^4],$$

where $n\mathbb{E}\left[\frac{1}{J_i^2}\right] = O(1)$. Therefore we have the results. For $c_{i,\text{HO}}$, the proof is similar.

3. (a) Follows from the finite fourth moment and the inequality below for any $s > 0$,

$$\Pr\left(|\sqrt{n}(\bar{\theta} - \mathbb{E}[\theta_i])| > s\right) \leq \frac{\mathbb{E}[\theta_i^2]}{s^2}. \quad (\text{Chebyshev's inequality})$$

(b) Follows from the property of $\bar{\theta}$ and the second property of 2:

$$\bar{X} = \bar{\theta} + \bar{\epsilon} = \mathbb{E}[\theta_i] + O_p(n^{-1/2}) + o_p(n^{-1/2}) = \mathbb{E}[\theta_i] + O_p(n^{-1/2}).$$

(c) By Chebyshev's inequality and (6),

$$\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^n\frac{1}{J_i}\hat{\sigma}_i^2 - \mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2\right]\right| > s\right) \leq \frac{\mathbb{E}\left[\frac{1}{J_i^2}\left(\frac{K\sigma_i^4}{J_i} + \sigma_i^4\right)\right]}{s^2} \rightarrow 0,$$

and also

$$\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2\right] \leq \sqrt{n\mathbb{E}\left[\frac{1}{J_i^2}\right]\mathbb{E}[\sigma_i^4]} = O(1).$$

Thus

$$\frac{1}{n}\sum_{i=1}^n\frac{1}{J_i}\hat{\sigma}_i^2 = O_p(n^{-1/2}).$$

4. (a) Since $u_i \perp \epsilon_{i,j} \mid \theta_i$, by Chebyshev's inequality,

$$\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^nc_i^k u_i\right| > s\right) \leq \frac{\mathbb{E}[c_i^{2k}u_i^2]}{s^2} \leq \frac{\sigma_u^2}{s^2}$$

Thus

$$\frac{1}{n}\sum_{i=1}^nc_i^k u_i = O_p(n^{-1/2}).$$

(b) By Chebyshev's inequality,

$$\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^nc_i^k u_i \bar{\epsilon}_i\right| > s\right) \leq \frac{\mathbb{E}[c_i^{2k}u_i^2\bar{\epsilon}_i^2]}{s^2} \leq \frac{\sigma_u^2}{s^2}\sqrt{\mathbb{E}\left[\frac{1}{J_i^4}\right]\mathbb{E}[\sigma_i^8]} \rightarrow 0.$$

Thus

$$\frac{1}{n}\sum_{i=1}^nc_i^k u_i \bar{\epsilon}_i = o_p(n^{-1/2}).$$

For $c_{i,\text{HO}}$, the proof is similar.

□

Lemma D.2. *Under Assumption D.1, Assumption D.2, we have*

$$\frac{1}{n} \sum_{i=1}^n c_i^2 \bar{\epsilon}_i^2 = \frac{1}{n} \sum_{i=1}^n c_i^2 \frac{1}{J_i} \hat{\sigma}_i^2 + o_p(n^{-1/2}).$$

For $c_{i,\text{HO}}$, the result is similar.

Proof of Lemma D.2. First, notice that the difference

$$\begin{aligned} \bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 &= \bar{\epsilon}_i^2 - \frac{1}{J_i(J_i-1)} \sum_{j=1}^{J_i} \epsilon_{i,j}^2 + \frac{1}{J_i-1} \bar{\epsilon}_i^2 \\ &= \frac{1}{J_i(J_i-1)} \left(\sum_{j=1}^{J_i} \epsilon_{i,j} \right)^2 - \frac{1}{J_i(J_i-1)} \sum_{j=1}^{J_i} \epsilon_{i,j}^2 \\ &= \frac{2}{J_i(J_i-1)} \sum_{k \leq j} \epsilon_{i,k} \epsilon_{i,j}. \end{aligned}$$

Then the properties of the difference are

$$\begin{aligned} \mathbb{E} \left[\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right] &= 0, \\ \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \mid J_i, \sigma_i^2 \right] &= \frac{2\sigma_i^4}{J_i(J_i-1)}. \end{aligned}$$

For c_i^2 , by the mean value theorem,

$$\begin{aligned} c_i^2 &= \left(\frac{V}{V + \frac{1}{J_i} \hat{\sigma}_i^2} \right)^2 \\ &= 1 - 2 \frac{1}{V J_i} \hat{\sigma}_i^2 + 3 \frac{1}{(1 + \omega_i)^4} \frac{1}{V^2 J_i^2} (\hat{\sigma}_i^2)^2, \end{aligned}$$

where ω_i is between 0 and $\frac{1}{V J_i} \hat{\sigma}_i^2$. Therefore,

$$\begin{aligned}
& \left| \sqrt{n} \mathbb{E} \left[c_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\
& \leq \left| \sqrt{n} \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| + \frac{2}{V} \left| \sqrt{n} \mathbb{E} \left[\frac{1}{J_i} \hat{\sigma}_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\
& \quad + \frac{3}{V^2} \left| \sqrt{n} \mathbb{E} \left[\frac{1}{(1 + \omega_i)^4} \frac{1}{J_i^2} (\hat{\sigma}_i^2)^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| \\
& \leq 0 + \frac{2}{V} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^2} (\hat{\sigma}_i^2)^2 \right] \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]} \\
& \quad + \frac{3}{V^2} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right] \mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]} \tag{Cauchy-Schwarz} \\
& \leq \frac{2}{V} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K}{J_i} \sigma_i^4 + \sigma_i^4 \right) \right]} \sqrt{n \mathbb{E} \left[\frac{2\sigma_i^4}{J_i (J_i - 1)} \right]} \\
& \quad + \frac{3}{V^2} \sqrt{\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right]} \sqrt{n \mathbb{E} \left[\frac{2\sigma_i^4}{J_i (J_i - 1)} \right]}. \\
& \leq \frac{2}{V} \sqrt{\sqrt{n \mathbb{E} \left[\frac{1}{J_i^6} \right] \mathbb{E} [K^2 \sigma_i^8] + n \mathbb{E} \left[\frac{1}{J_i^4} \right] \mathbb{E} [\sigma_i^8]} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^2 (J_i - 1)^2} \right] \mathbb{E} [4\sigma_i^8]} \\
& \quad + \frac{3}{V^2} \sqrt{\sqrt{n \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right]} \sqrt{n \mathbb{E} \left[\frac{1}{J_i^2 (J_i - 1)^2} \right] \mathbb{E} [4\sigma_i^8]}. \tag{Cauchy-Schwarz}
\end{aligned}$$

Since we have Assumption [D.2](#) and that

$$\sqrt{n} \mathbb{E} \left[\frac{1}{J_i^4} (\hat{\sigma}_i^2)^4 \right] \rightarrow 0,$$

from Lemma [D.1.1e](#), the above converges to 0. Thus, the expectation

$$\sqrt{n} \mathbb{E} \left[c_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \rightarrow 0.$$

Next, we show the convergence to expectations by Chebyshev's inequality:

$$\begin{aligned} & \Pr \left(\sqrt{n} \left| \frac{1}{n} \sum_{i=1}^n c_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) - \mathbb{E} \left[c_i^2 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right) \right] \right| > s \right) \\ & \leq \frac{\mathbb{E} \left[c_i^4 \left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]}{s^2} \leq \frac{\mathbb{E} \left[\left(\bar{\epsilon}_i^2 - \frac{1}{J_i} \hat{\sigma}_i^2 \right)^2 \right]}{s^2} = \frac{\mathbb{E} \left[\frac{2\sigma_i^4}{J_i(J_i-1)} \right]}{s^2} \rightarrow 0. \end{aligned}$$

And the proof for c_i is complete.

For $c_{i,\text{HO}}$, the proof is similar. □

Lemma D.3. *Under Assumption D.1, Assumption D.2, we have*

$$\hat{V} = V + O_p(n^{-1/2}).$$

Proof of Lemma D.3. By Lemma D.1,

$$\begin{aligned} \hat{V} &= \frac{1}{n} \sum_{i=1}^n (\bar{X}_i - \bar{X})^2 - \frac{n-1}{n^2} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 \\ &= \frac{1}{n} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{n} \sum_{i=1}^n \bar{\epsilon}_i^2 - (\bar{\theta} - \mathbb{E}[\theta_i])^2 - \bar{\epsilon}^2 + \frac{2}{n} \sum_{i=1}^n (\theta_i - \mathbb{E}[\theta_i]) \bar{\epsilon}_i \\ &\quad - 2(\bar{\theta} - \mathbb{E}[\theta_i]) \bar{\epsilon} - \frac{n-1}{n^2} \sum_{i=1}^n \frac{1}{J_i} \hat{\sigma}_i^2 \\ &= V + O_p(n^{-1/2}) + O_p(n^{-1/2}) - O_p(n^{-1}) - o_p(n^{-1}) + o_p(n^{-1/2}) \\ &\quad - o_p(n^{-1}) - O_p(n^{-1/2}) \\ &= V + O_p(n^{-1/2}). \end{aligned}$$

□

Lemma D.4. *Under Assumption D.1, Assumption D.2, we have*

$$\hat{V}_{HO} = V + O_p(n^{-1/2}).$$

Proof of Lemma D.4. By conventional law of large number and central limit theorem

arguments,

$$\begin{aligned}\hat{V}_{\text{HO}} &= \frac{1}{n} \sum_{i=1}^n (\bar{X}_{1,i} - \bar{X}_1) (\bar{X}_{2,i} - \bar{X}_2) \\ &= V + O_p(n^{-1/2})\end{aligned}$$

□

Proposition D.1. *Suppose the asymptotic framework satisfies Assumption D.2. Then under Assumption D.1 we have*

$$\sqrt{n} (\hat{\beta} - \beta) \rightarrow_d N \left(0, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{V^2} \right).$$

Proof of Proposition D.1. Based on a similar set of lemmas to Appendix B, the claim follows from the same arguments as in Theorem 3.3. □

Proposition D.2 (Proposition 3.2). *Suppose the asymptotic framework satisfies Assumption D.2. Then under Assumption D.1, there exist cases where*

$$\sqrt{n} (\hat{\beta}_{\text{HO}} - \beta) \rightarrow_d N \left(\frac{\beta}{\text{Var}(\theta_i)}, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{(\text{Var}(\theta_i))^2} \right).$$

Proof of Proposition D.2. For the shrinkage weight

$$c_{i,\text{HO}} = \frac{V}{\frac{1}{J_i} \mathbb{E} [\sigma_i^2] + V},$$

we first show properties of the regression coefficients $\hat{\beta}_{\text{HO},c}$, and then show that $\sqrt{n} (\hat{\beta}_{\text{HO}} - \beta) \rightarrow_p \sqrt{n} (\hat{\beta}_{\text{HO},c} - \beta)$.

For $\hat{\beta}_{\text{HO},c}$, based on a similar set of lemmas to Appendix B, from the same derivation as the proof of Lemma 3.1, we have

$$\sqrt{n} (\hat{\beta}_{\text{HO},c} - \beta) = \frac{\beta \sqrt{n} T_{1,n} - \beta \sqrt{n} T_{2,n} + \sqrt{n} T_{3,n}}{T_{2,n}},$$

where the denominator

$$T_{2,n} = V + O_p(n^{-1/2}),$$

and the numerator

$$\begin{aligned}
& \beta\sqrt{n}T_{1,n} - \beta\sqrt{n}T_{2,n} + \sqrt{n}T_{3,n} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\beta c_{i,\text{HO}} (\theta_i - \mathbb{E}[\theta_i])^2 - \beta c_{i,\text{HO}}^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{J_i} \hat{\sigma}_i^2 \right] + c_{i,\text{HO}} (\theta_i - \mathbb{E}[\theta_i]) u_i \right] + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\beta c_{i,\text{HO}} (\theta_i - \mathbb{E}[\theta_i])^2 - \beta c_{i,\text{HO}}^2 \left[(\theta_i - \mathbb{E}[\theta_i])^2 + \frac{1}{J_i} \mathbb{E}[\sigma_i^2] \right] + c_{i,\text{HO}} (\theta_i - \mathbb{E}[\theta_i]) u_i \right] \\
&\quad - \beta \frac{1}{n} \sum_{i=1}^n c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E}[\sigma_i^2]) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i - \beta \frac{1}{n} \sum_{i=1}^n c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E}[\sigma_i^2]) + o_p(1).
\end{aligned}$$

For ξ_i , similar to Lemma 3.1, we have $\mathbb{E}[\xi_i] = 0$, $\mathbb{E}[\xi_i^2] \rightarrow \mathbb{E}[u_i^2 (\theta_i - \mathbb{E}[\theta_i])^2]$. However, the second term of the numerator might not disappear.

Next, following from the same arguments as Theorem 3.3, we have $\sqrt{n}(\hat{\beta}_{\text{HO}} - \beta) \rightarrow_p \sqrt{n}(\hat{\beta}_{\text{HO},c} - \beta)$.

Now, focusing on the bias term, by Chebyshev's inequality and (6), for any $s > 0$,

$$\begin{aligned}
& \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E}[\sigma_i^2]) - \mathbb{E} \left[c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E}[\sigma_i^2]) \right] \right| > s \right) \\
&\leq \frac{\mathbb{E} \left[\frac{1}{J_i^2} (\hat{\sigma}_i^2 - \mathbb{E}[\sigma_i^2])^2 \right]}{s^2} \\
&\leq \frac{\mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K\sigma_i^4}{J_i} + (\sigma_i^2 - \mathbb{E}[\sigma_i^2])^2 \right) \right]}{s^2} \rightarrow 0.
\end{aligned}$$

Suppose with equal probabilities, $\sigma_i^2 = 12\gamma V$, $J_i = \lfloor 2\sqrt{n} \rfloor$ or $\sigma_i^2 = 8\gamma V$, $J_i = \lfloor \frac{2}{3}\sqrt{n} \rfloor$, $\gamma > 0$. Then we have

$$n\mathbb{E} \left[\frac{1}{J_i^2} \right] \sim \frac{1}{2}n * \frac{1}{(2\sqrt{n})^2} + \frac{1}{2}n * \frac{1}{(\frac{2}{3}\sqrt{n})^2} \rightarrow \frac{5}{4},$$

$$\begin{aligned}
& \mathbb{E} \left[c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2]) \right] \\
&= \mathbb{E} \left[c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\sigma_i^2 - \mathbb{E} [\sigma_i^2]) \right] \\
&\sim \frac{2}{2} \left(\frac{V}{\frac{10}{2\sqrt{n}} + V} \frac{\sqrt{n}}{2\sqrt{n}} - \frac{V}{\frac{10}{\frac{2}{3}\sqrt{n}} + V} \frac{\sqrt{n}}{\frac{2}{3}\sqrt{n}} \right) \gamma V \\
&\rightarrow -\gamma V.
\end{aligned}$$

Then by the continuous mapping theorem, in this case we have

$$\sqrt{n} (\hat{\beta}_{\text{HO}} - \beta) \rightarrow_d N \left(\gamma\beta, \frac{\mathbb{E} [u_i^2 (\theta_i - \mathbb{E} [\theta_i])^2]}{V^2} \right),$$

which proves the first result of the theorem.

To prove the second result of the theorem, it suffices to show that if σ_i^2 and J_i are independent, then

$$\frac{1}{n} \sum_{i=1}^n c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2]) = o_p(1).$$

Since $c_{i,\text{HO}}$ only depends on J_i , by the independence, the expectation

$$\mathbb{E} \left[c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2]) \right] = 0.$$

Then, by Chebyshev's inequality,

$$\begin{aligned}
& \Pr \left(\left| \frac{1}{n} \sum_{i=1}^n c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2]) - \mathbb{E} \left[c_{i,\text{HO}}^2 \frac{\sqrt{n}}{J_i} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2]) \right] \right| > s \right) \\
&\leq \frac{\mathbb{E} \left[c_{i,\text{HO}}^4 \frac{1}{J_i^2} (\hat{\sigma}_i^2 - \mathbb{E} [\sigma_i^2])^2 \right]}{s^2} \\
&\leq \frac{\mathbb{E} \left[\frac{1}{J_i^2} \left(\frac{K\sigma_i^4}{J_i} + (\sigma_i^2 - \mathbb{E} [\sigma_i^2])^2 \right) \right]}{s^2} \rightarrow 0,
\end{aligned}$$

which proves the second result of the theorem. \square