Automatic Inference for Value-Added Regressions*

Tian Xie

Department of Economics, University College London

November 10, 2025

Click here for the latest version

Abstract

A large empirical literature performs regressions of outcomes on empirical Bayes shrinkage value-added estimates, yet little is known about whether this approach leads to valid inference. We study a general class of value-added estimators and the properties of the resulting regression coefficients. We show that inference can be invalid if the shrinkage estimator does not account for heteroskedasticity in the underlying noise. 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:` shrinkage estimators, teacher value-added, error in variables

`JEL classification codes:` C12.

*Email address: tian.xie.20@ucl.ac.uk. 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, Eric Mbakop, Matthias Parey, Kirill Ponomarev, Silvia Sarpietro, Michela Tincani, Edward Vytlacil, Guanyi Wang, Weining Wang, Andrei Zeleneev, and participants of IAAE Conference (2025), Bristol Econometric Study Group (2025), The 2nd UCL-CeMMAP Ph.D. Econometrics Research Day, Midwest Econometrics Group Conference (2025), and seminars at UCL and Yale University. All errors are my own.

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. In many of the settings, the shrinkage estimates are 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 individual effects.

It is common for researchers to treat the shrinkage estimates as if they were true individual effects and conduct conventional inference in the downstream regression. This approach implicitly assumes the property of automatic inference, i.e., no adjustment is required to achieve nominal coverage rates of confidence intervals. However, the shrinkage estimates generally differ from the true effects, introducing measurement-error concerns. Moreover, they are constructed as first-stage estimates, raising the possibility of generated-regressor issues (Pagan, 1984). Thus, there may be threats to the validity of both the estimator (via measurement error bias), and the inference (via incorrect standard errors).

In empirical applications, various linear shrinkage estimators have been used as regressors, complicating the properties of the resulting downstream regression. A leading example is the literature studying how teachers' value-added for test scores impacts students' future outcomes. Based on students' test scores as measurements, shrinkage estimates 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. Similar empirical work 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). This 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 the validity and efficiency of downstream inference remains unclear.

This paper is the first to formalize such 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 demonstrates that automatic inference hinges on the treatment of heteroskedastic measurement error. We show that failing to account for heteroskedasticity in the shrinkage weights leads to *invalid* downstream inference, with 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 efficient, being asymptotically equivalent to the infeasible OLS regression on the true latent effects.

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 are improper. It introduces a non-classical measurement error problem, where the bias can be amplification rather than attenuation. Inference remains valid when the two sources of heterogeneity are independent. Incorporating both sources ensures robustness to such dependence and achieves efficiency 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. 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 specifications 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 coefficient estimates are asymptotically unbiased and inference is valid. We show that for some scenarios this is the case, while for others it is not.

To formalize our analysis, we present results for baseline methods and for individual-weight shrinkage estimators. We first restate the established baseline result that the unshrunk individual mean suffers from classical errors-in-variables attenuation bias and yields invalid inference when used as a downstream regressor. We then show our main results on individual-weight shrinkage. First, we delineate 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 efficiency and inferential validity. 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. It provides a useful benchmark for our primary results on the more empirically prevalent individual-weight approach. While our focus as been on individual-weight shrinkage that assumes (as is standard) independence between individual effects and variance of measurement error, recent work has derived empirical Bayes estimators allowing such dependence. In on-going work, we extend our analysis to this new class of estimators and show that they also yield valid downstream inference.

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 is the method using the firm discrimination data of Kline et al. (2022) and the second is 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 support 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 when used as regressors, and it reaffirms the positive effect of school value-added on private school fees.

Related Literature This paper relates to the literature that focuses on common-weight shrinkage, which is discussed in the earliest work to use shrinkage estimators as downstream regressors. Regressing on common-weight shrinkage estimators can be traced back to Whittemore (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 shrinkage 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 study 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 the simple shrinkage strategies. Beyond such consistency results, there is no general framework investigating the validity of inference and standard errors for individual-weight downstream regressions. 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, this category includes common-weight shrinkage given its equivalence to such methods. Kline, Saggio, and Sølvsten (2020) and Bonhomme and Denis (2024) emphasize correct estimation of moments for value-added that generates correct regression estimates as a component. 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 asymptotic normality, consistency of standard errors, and discusses the validity of standard inference approaches. It also provides some practical implementation guidelines. Section 4 reports Monte Carlo simulations illustrating the finite-sample properties of the estimators. These results reinforce out main findings that properly constructed individualized shrinkage estimators yield valid inference, while improperly constructed ones can fail. Section 5 revisits the study of Kline et al. (2022) to labor market discrimination. Section 6 revisits the study of Andrabi et al. (2025) to school value-added. Finally, Section 7 concludes. An appendix contains the additional lemmas, proofs, and extensions.

2 Setup

2.1 Model

Each unit i is associated with a latent individual effect θ_i, which we will refer to in what follows as value-added. Let $\boldsymbol{\theta} \coloneqq (\theta_i)_{i=1}^n$ 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, $\mathbb{E}[u_i\theta_i] = 0$. One may be interested in how the effect on short-term test scores translates to long-term outcomes—such as college attendance or earnings—through the regression 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 $(Y_i, X_i) \coloneqq (Y_i, X_{i,1}, ..., X_{i,J_i})$, $i = 1, ..., n$, where Y_i is the outcome, $X_{i,j}$ is the measurement of θ_i, and J_i is the numebr 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

$$X_{i,j} = \theta_i + \epsilon_{i,j}, \tag{2}$$

where $\epsilon_{i,j}$ is the noise term. Within the example, $X_{i,j}$ is the short-term test score outcome of student j taught by teacher i. The individual mean score is denoted by $\bar{X}_i$ and the pooled mean score by $\bar{X}$. In this paper, we primarily focus on the case where the noise is independent of the value-added, i.e. $\epsilon_{i,j} \perp\!\!\!\perp \theta_i$. We also assume the noise is independent of the structural error u_i, i.e. $\epsilon_{i,j} \perp\!\!\!\perp u_i$. Before proceeding, we discuss how the observation-specific sample size J_i relates to n.

2.2 Asymptotic Framework

As is standard, empirical Bayes shrinkage estimators of $\boldsymbol{\theta}$ are used as the regressor. It is commonly assumed that $\bar{X}_i \mid \theta_i \sim N(\theta_i, \sigma_i^2/J_i)$ in empirical Bayes, where σ_i^2 is the variance of noise $\epsilon_{i,j}$. Without distributional knowledge about the noise, the normality of $\bar{X}_i$ is typically justified by the central limit theorem with J_i reasonably large (Walters, 2024). It is therefore logically consistent to adopt a framework where both J_i and n are "large" in a suitable sense. In the teacher value-added example, J_i often has a similar magnitude to $\sqrt{n}$, with the number of teachers in the thousands and the number of students per teacher in the tens. For example, 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 is 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. The measurement error in θ_i is proportional to J_i^{-1} whereas the sampling error for estimation of β in (1) is proportional to $n^{-1/2}$. We therefore adopt a framework so that both are of a similar magnitude. In finite samples, both sources of measurement error are present. The framework we adopt ensures both sources are present asymptotically, so that the asymptotic distributions we derive give useful approximations to the finite-sample problem faced by the researcher.

To formalize the idea, we consider the following asymptotic framework. As $n \to \infty$, the DGP of the other variables are fixed, but for J_i we shall assume

$$\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\right] \to \kappa \in [0, +\infty), \tag{3}$$

$$\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\right] \to 0. \tag{4}$$

Note that if we have the same number of measurements, $J_i = J$, then (4) is implied by (3). Thus (3) is the essential condition, under which the asymptotic problem minics the finite sample problem. The product $\sqrt{n}\mathbb{E}\left[J_i^{-1}\right]$ quantifies the relative scale of measurement error 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] A finite positive κ thus seems appropriate for the context in which tens of students and thousands of teachers coexist.

Our asymptotic framework is s related to the small-variance approximation in Chesher (1991) and Evdokimov and Zeleneev (2019, 2023). 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).

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 work: 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. Our interest is on the validity of inference in Step 2 when different shrinkage estimators are used in Step 1.

The primary implication of our framework is that the measurement error in $\bar{X}_i$ does not vanish relative to the sampling error for β. 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 the necessary baseline and motivation for the other estimators.

We consider four primary classes of estimators for $\hat{\boldsymbol{\theta}}$, 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: homoskedastic individual-weight shrinkage (HO), which models heterogeneity in measurement preci-

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

sion solely by the number of measurements (J_i), and heteroskedastic individual-weight shrinkage (HE), which incorporates both sources of heterogeneity by also modeling idiosyncratic noise variances. We study homoskedastic individual-weight shrinkage in Section 3.2 and heteroskedastic individual-weight shrinkage in Section 3.3. Finally, as a benchmark from equal shrinkage, 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, unit value-added for unit i is simply estimated by the individual average: $\hat{\theta}_{i,\text{FE}} = \bar{X}_i$, for $i = 1, ..., n$.

As we show formally below, using $\bar{X}_i$ as a regressor suffers from the classical errors-in-variable (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. Before formally presenting the result, 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. $\mathbb{E}[\epsilon_{i,j} \mid \sigma_i^2] = 0$, $\mathbb{E}\left[\epsilon_{i,j}^2 \mid \sigma_i^2\right] = \sigma_i^2$, $\mathbb{E}\left[|\epsilon_{i,j}|^L \mid \sigma_i^2\right] \leq K\sigma_i^L$, *for some* $L \geq 3$. *Also,* $\epsilon_{i,j} \perp\!\!\!\perp \theta_i$.
4. $u_i \perp\!\!\!\perp \epsilon_{i,j} \big| \theta_i$.
5. $\mathbb{E}[\theta_i^4] < \infty$.
6. $\mathbb{E}[\sigma_i^{16}] < \infty$.
7. $\sqrt{n}\mathbb{E}\left[J_i^{-1}\right] \to \kappa$, $\kappa \in [0, +\infty)$.
8. $\sqrt{n}\mathbb{E}\left[J_i^{-2}\right] \to 0$.

Assumption 3.1.1 assumes independent numbers of measurements J_i. In Appendix D, the condition will be relaxed and 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 the 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 $\epsilon_{i,j} \perp\!\!\!\perp \theta_i$ is standard in empirical Bayes literature to justify shrinkage estimators without covariates. Assumption 3.1.4 excludes further effect from measurement errors to 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 on J_i as discussed in Section 2.

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

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

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

The proposition's result holds even without Assumption 3.1.8. In the special case where $\kappa = 0$ (meaning measurement error is asymptotically negligible), the bias term disappears, the asymptotic distribution of $\sqrt{n}\left(\hat{\beta}_{\text{FE}} - \beta\right)$ is centered at zero, and standard inference would be valid. 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. However, its asymptotic distribution is centered away from zero due to the 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, $\operatorname{Var}\left(\bar{X}_i \mid \theta_i\right) = \sigma^2/J_i$. We refer to this case 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 may take this heterogeneity into account or not.

Following the estimators proposed by Jacob and Lefgren (2007) and Kane and Staiger (2008), the homoskedastic individual-weight shrinkage estimator $\boldsymbol{\theta}$ is constructed as

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

The estimator is the empirical Bayes posterior mean under normality of θ_i and $\epsilon_{i,j}$. The weight is $w_{i,\mathrm{HO}}$, obtained by replacing the oracle shrinkage weight

$$\frac{\operatorname{Var}(\theta_i)}{\sigma^2/J_i + \operatorname{Var}(\theta_i)}$$

with its empirical counterparts. These variances are estimated as

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

where $\bar{X}_{i,t}$, $\bar{X}_{i,t-1}$ are from splitting the measurements for each i into two subsets and computing their averages. They can also be averages for two time periods in a different context. 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 $\hat{\beta}_{\mathrm{HO}}$ of regressing Y_i on $\hat{\theta}_{i,\mathrm{HO}}$ 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, homoskedasticity entails that regressing Y_i on $\hat{\theta}_{i,\mathrm{HO}}$ yields asymptotic unbiasedness and valid inference on β.

We then derive results for the heteroskedastic case. Here we show that if the true DGP is heteroskedastic, there would be problems in inference when homoskedastic weights are applied. Denote the variance of $\bar{X}_i \mid \theta_i$ by σ_i^2/J_i. Regressing Y_i on $\hat{\theta}_{i,\mathrm{HO}}$ 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 distorts 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*

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

If instead J_i and σ_i^2 are independent, then

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

Proposition 3.2 clarifies the conditions required for regressing Y_i on $\hat{\theta}_{i,\mathrm{HO}}$ 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 depends on the dependence of J_i and σ_i^2. Thus, the bias can 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 directly addresses the sensitivity of $\hat{\theta}_{i,\mathrm{HO}}$.

The heteroskedastic individual-weight shrinkage estimator is defined as

$$\begin{aligned}
\hat{\theta}_{i,\mathrm{HE}} &\coloneqq w_{i,\mathrm{HE}}\bar{X}_i + (1 - w_{i,\mathrm{HE}})\bar{X}, \text{ where} \\
w_{i,\mathrm{HE}} &\coloneqq \frac{\hat{V}}{\frac{1}{J_i}\hat{\sigma}_i^2 + \hat{V}}.
\end{aligned} \tag{5}$$

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

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

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

We take an intermediate step before deriving the properties of $\hat{\beta}_{\mathrm{HE}}$, the downstream OLS estimator from regressing Y_i on $\hat{\theta}_{i,\mathrm{HE}}$. The estimator $\hat{\theta}_{i,\mathrm{HE}}$ 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 $\hat{\beta}_{c,\mathrm{HE}}$ from regressing Y_i on the semi-oracle estimator $\hat{\theta}_{i,c,\mathrm{HE}}$ (with prior mean still estimated from the data),

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

Lemma B.3 shows that the estimated variance $\hat{V}$ is a consistent estimator for the true variance $\mathrm{Var}\,(\theta_i)$. Thus, we build the properties of $\hat{\beta}_{\mathrm{HE}}$ based on those of $\hat{\beta}_{c,\mathrm{HE}}$. The following result shows that the OLS estimator $\hat{\beta}_{c,\mathrm{HE}}$ from regression Y_i on $\hat{\theta}_{i,c,\mathrm{HE}}$ 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*

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

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

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

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

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

and so

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

Theorem 3.3 establishes that the HE estimator $\hat{\theta}_{i,\mathrm{HE}}$ performs first-order asymptotically equivalent to the $\hat{\theta}_{i,c,\mathrm{HE}}$ as the regressor. Moreover, there is no efficiency loss relative to the infeasible regression of Y_i on the latent θ_i (See the unconditional moment case of Chamberlain, 1987). The asymptotic distribution doesn't depend on the value of κ. Thus, the OLS regression estimator $\hat{\beta}_{\mathrm{HE}}$ is robust to the amount of measurement error (under the conditions of Theorem 3.3).

Recall from Proposition 3.1 that the asymptotic distribution of the estimator from regressing on unshrunk $\hat{\theta}_{i,\mathrm{FE}}$ 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 for the EIV problem, thereby resolving the inferential issue in the baseline FE case.

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 $\hat{\theta}_{i,\text{HE}}$. This enables implementation of inference in practice.

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

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

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

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:*

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

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:

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

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.

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 θ_i. 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, $\hat{\boldsymbol{\theta}}_{\text{HE}}$ is known to achieve a lower mean squared error (MSE) 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:

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

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 $\hat{\theta}_{i,\mathrm{CW}}$ leads to an OLS estimator of β given by

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

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

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

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, then we can construct an IV estimator:

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

The IV estimator is adjusting the estimator of regressing Y_i on $\bar{X}_{i,1}$ by a factor. With a proper adjustment factor, we can exploit more information from regressing Y_i on $\bar{X}_i$ and achieve better estimation than using the subset $\bar{X}_{i,1}$.

Because of its connection to the bias correction method and the IV estimator, common-weight shrinkage can achieve as good performance as those methods for β, requiring as weak assumptions as them. Even though it ignores the heterogeneous signal-to-noise ratio and pool every individual at a uniform ratio, common-weight shrinkage provides valid inference on β. The following result formalizes this point. It is proved in Appendix B.

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

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

Therefore, the CW estimator also addresses the inferential problem from the FE baseline. Its primary distinction from the HE estimator is how it achieves this: the CW approach applies a uniform correction factor (equivalent to bias-correction or IV), while the HE estimator models unit-level heterogeneity. Despite these different mechanisms, both estimators are first-order asymptotically equivalent and achieve the same efficient asymptotic distribution.

4 Simulations

4.1 Simulation Design

In the first set of simulations, we focus on the comparison of regressing on the shrinkage estimator $\hat{\theta}_{i,\mathrm{HE}}$, the sample mean $\bar{X}_i$, i.e. $\hat{\theta}_{i,\mathrm{FE}}$, and the true latent θ_i. In each of the $S = 3000$ simulations, we generate the data, run the regression of Y_i on generated 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 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 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 general distributions of measurement error are allowed as long as the moment conditions are satisfied.

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

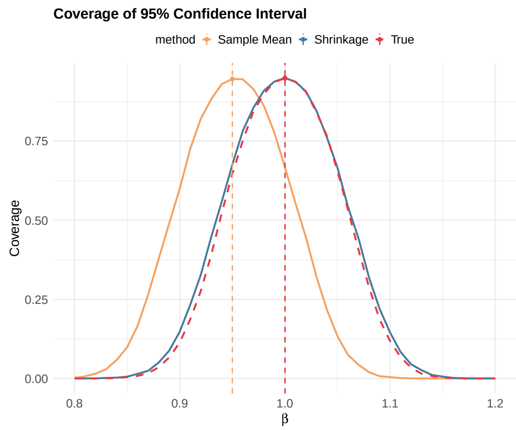
In the third set of simulations, we again compute the coverage rate for each method. We check the performance with random J_i drawn from Poisson(20), and with the sample size $n = 1000$. Here $\hat{\kappa}$ is approximately 1.58 but larger due to convexity. With other parameters unchanged, now we can have correlations 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 conditions in Assumption D.1 satisfied.

4.2 Results

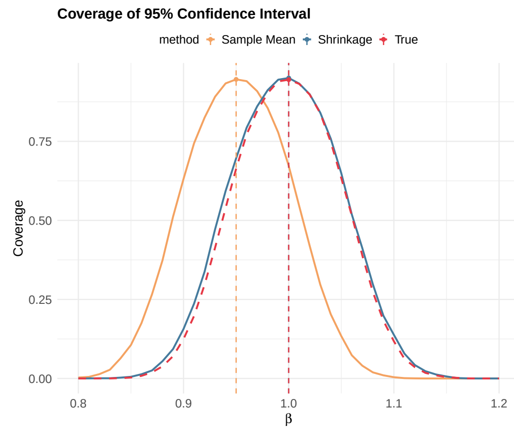
The first set of coverage results are presented in Figure 1. Coverage for regressing on HE (the blue line) performs well in both normal and non-normal settings. The coverage rates are close to 95% at the true β, and the curve is always very close to the infeasible case (the dashed red line) of regressing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

The second set of results are reported in Table 1. The results indicate that HE has a smaller MSE of β, higher coverage rate, smaller bias, and smaller MSE of $\boldsymbol{\theta}$ than CW when the ratio $\sqrt{n}\mathbb{E}\left[J_i^{-1}\right]$ is reasonable. When n and J tend to be large, the performance of CW and HE are close, but HE always dominates and is even more preferable when the sample size is relatively small. As discussed in Section 3.4, we know that $\hat{\beta}_{\mathrm{CW}}$ is first-order asymptotically equivalent to $\hat{\beta}_{\mathrm{HE}}$. Combined with the simulation results, we can see that 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 the proposed method (the blue line) and HO (the purple line). We can see that the proposed method works well in both settings, close to the infeasible case (the dashed red line). Instead, HO only works well in the first setting. When J_i and σ_i^2 are correlated, $\hat{\beta}_{\mathrm{HO}}$ is biased, reflected by a shift to the right in the coverage plot



(a) Normal Noise



(b) Gamma Noise

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 proposed method of regressing on the shrinkage estimator $\hat{\theta}_i$ (HE). The orange line represents regressing on the sample mean $\bar{X}_i$ (FE).

Figure 1: Coverage Rates Under Different Noise Distributions

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

5 Empirical Application: Firm Discrimination

In this section, we use the HE method in the context of Kline et al. (2022) about the extent to which large U.S. employers systemically discriminate job applicants based on race. Their study utilizes correspondence audits, where fictitious resumes with

Table 1: Method Comparison Across Scenarios

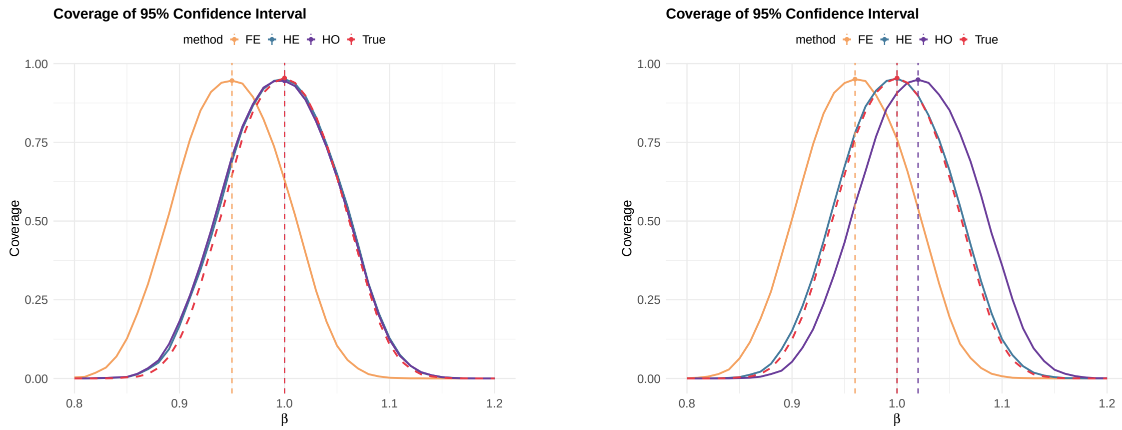
Scenario	Method	$\sqrt{n \times \text{MSE}(\beta)}$	Coverage Rate (%)	Bias	MSE($\boldsymbol{\theta}$)
n=50, J=10	True θ_i	1.030	94.73	0.116	0
$\sqrt{n}/J \approx 0.71$	HE	1.497	91.90	0.166	0.322
	CW	2.335	87.83	0.225	0.374
	FE	2.690	27.67	0.354	0.552
n=225, J=10	True θ_i	1.016	94.17	0.054	0
$\sqrt{n}/J \approx 1.5$	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	True θ_i	1.009	95.13	0.025	0
$\sqrt{n}/J \approx 1.58$	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 $\boldsymbol{\theta}$ measures the estimation error for individual effects. A larger $\sqrt{n}/J$ indicates a higher ratio of measurement error to sampling error.

randomized 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) on 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



(a) $J_i \perp\!\!\!\perp \sigma_i^2$ (b) J_i correlated with σ_i^2

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 $\hat{\theta}_{i,\text{HE}}$ (HE). The purple line represents the regressing on the homoskedastic individual-weight shrinkage estimator $\hat{\theta}_{i,\text{HO}}$ (HO). The orange line represents regressing on the sample mean $\bar{X}_i$ (FE).

Figure 2: Coverage Rates Under Different Dependence

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 number of contacts and number of received applications) for White and that for Black applications. We follow the model similar to Kline et al. (2022), where the racial contact gap is given by

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

with normality arising from the central limit theorem approximation.

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. We compare the performance of our method with the fixed effects (FE) estimator 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, along with lower mean squared error (MSE). Despite a small $\sqrt{n}/J$, FE still exhibits attenuation bias. In terms of inference, the proposed method robustly rejects the null hypothesis of no predictive effect, whereas FE fails to reject this null hypothesis for waves 1 and 2. These findings demonstrate that the proposed estimator improves both the accuracy of estimation and the validity of inference for the regression coefficient.

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

	$\bar{X}_i$				$\hat{\theta}_{i,\text{HE}}$			
Wave	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
MSE(β)	0.630				0.474			

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.

6 Application: School Value-Added in Pakistan

In this section, we use the HE method in the context of Andrabi et al. (2025) about estimating the school value-added in Pakistan.

6.1 Data and Empirical Setting

The analysis utilizes the LEAPS project dataset, a rich longitudinal panel from rural Punjab, Pakistan. The data track 71,677 child-year test scores across more than 800 schools from 2003–2006, forming one of the largest such panels in a developing country. 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 $\mathbb{E}_n\left[J_i^{-1}\right] = 0.120$. Therefore $\hat{\kappa} = \sqrt{n}\mathbb{E}_n\left[J_i^{-1}\right] \approx 4.08$.

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 the original study. The second is HE, which is fully individualized by allowing 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 $20 \leq J_i \leq 80$, following the original paper). The tables compare the HE estimator (left panel) and the HO estimator (right panel) 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. For example, in the full-sample specification with controls (Table 3, Column 2), the HE point estimate is 793.013 and is statistically significant at the 1% level.

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

	(1)	(2)
	Dep. Var.: School Fees	
Empirical Bayes SVA	1054.128***	793.013***
	(298.150)	(264.052)
Mean Mother Education		2.314
		(123.829)
Mean Father Education		203.270
		(142.282)
Mean HH Asset Index		263.334***
		(46.927)
Adjusted R^2	0.23	0.29
Number of Observations	1154	1144
Number of Clusters	315	315

	(1)	(2)
	Dep. Var.: School Fees	
Empirical Bayes SVA	991.371***	719.858**
	(320.963)	(280.882)
Mean Mother Education		4.697
		(122.849)
Mean Father Education		218.651
		(142.686)
Mean HH Asset Index		263.074***
		(46.820)
Adjusted R^2	0.22	0.29
Number of Observations	1158	1148
Number of Clusters	318	318

Notes: Each column reports coefficients from regressions of private school fees on estimated school value-added. The left panel uses the HE estimator, and the right panel uses the HO estimator following Andrabi et al. (2025). Columns (2) add controls for mean mother education, mean father education, and mean household asset index. Standard errors (in parentheses) are clustered at the village level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

7 Conclusion

This paper investigates the inferential properties of individualized shrinkage estimators when used as downstream regressors, a common but not fully understood empirical practice. We formalize the conditions under which this plug-in approach yields valid downstream inference. Our central finding is that a correctly specified and fully individualized shrinkage estimator yields an asymptotically efficient estimator of the downstream coefficient. Crucially, we show that conventional OLS standard inference are asymptotically valid, justifying standard empirical practice. This result stands in contrast to the biased inference from the unshrunk FE baseline and the simpler individual-weight (HO) estimators that can fail when noise variance is correlated with the number of measurements. 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, 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 es-

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

	(1)	(2)
	Dep. Var.: School Fees	
Empirical Bayes SVA	1622.126***	1358.509***
	(454.324)	(391.156)
Mean Mother Education		-118.275
		(230.302)
Mean Father Education		393.320
		(271.010)
Mean HH Asset Index		359.615***
		(83.513)
Adjusted R^2	0.31	0.39
Number of Observations	592	591
Number of Clusters	150	150

	(1)	(2)
	Dep. Var.: School Fees	
Empirical Bayes SVA	1569.700***	1320.180***
	(451.274)	(384.742)
Mean Mother Education		-119.644
		(230.177)
Mean Father Education		402.032
		(270.856)
Mean HH Asset Index		360.131***
		(83.527)
Adjusted R^2	0.31	0.39
Number of Observations	592	591
Number of Clusters	150	150

Notes: The sample is restricted to schools with between 20 and 80 students, following Andrabi et al. (2025). The left panel reports the HE estimator, and the right panel replicates the HO estimates from the original study. Standard errors (in parentheses) are clustered at the village level. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

timator. For applied researchers, our results provide a clear theoretical 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. The regularity assumptions are Assumption 3.1. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{\text{FE}}$ as $\hat{\beta}$, and $\text{Var}(\theta_i)$ as V.

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

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,

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

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

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

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

Therefore, the numerator is

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

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

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

and also,

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

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

For ξ_i, since

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

$$\begin{aligned}
\mathbb{E}\left[\xi_i^2\right] &= \mathbb{E}\left[u_i^2\left(\theta_i - \mathbb{E}\left[\theta_i\right]\right)^2\right] + \beta^2\mathbb{E}\left[\left(\frac{1}{J_i}\sigma_i^2\right)^2\right] + \beta^2\mathbb{E}\left[\bar{\epsilon}_i^4\right] - 2\beta^2\mathbb{E}\left[\frac{1}{J_i}\sigma_i^2\bar{\epsilon}_i^2\right] \\
&= \mathbb{E}\left[u_i^2\left(\theta_i - \mathbb{E}\left[\theta_i\right]\right)^2\right] + o(1) - 2\beta^2\mathbb{E}\left[\frac{1}{J_i^2}\sigma_i^4\right] \\
&= \mathbb{E}\left[u_i^2\left(\theta_i - \mathbb{E}\left[\theta_i\right]\right)^2\right] + o(1), \qquad (6)
\end{aligned}$$

where the last line follows from Lemma B.1.

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

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

which is derived from (6) and the dominated convergence theorem, then we have

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

By (6) and Slutsky's theorem,

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

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

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

□

A.2 Proofs for Asymptotic Normality and Inference

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

We have

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

Respectively, by properties from Lemma B.1,

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

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

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

Therefore, for the denominator we have

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

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

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

For ξ_i, since

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

$$\begin{aligned}
\mathbb{E}\left[\xi_i^2\right] &= \beta^2\mathbb{E}\left[\left(c_i - c_i^2\right)^2\right]\mathbb{E}\left[\left(\theta_i - \mathbb{E}\left(\theta_i\right)\right)^4\right] + \beta^2\mathbb{E}\left[\frac{1}{J_i^2}c_i^4\hat{\sigma}_i^4\right] \\
&\quad + \mathbb{E}\left[c_i^2\right]\mathbb{E}\left[u_i^2\left(\theta - \mathbb{E}\left[\theta_i\right]\right)^2\right] \\
&\quad - 2\beta^2 V\mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2 c_i^2\left(c_i - c_i^2\right)\right] \\
&= o(1) + \beta^2\mathbb{E}\left[c_i^4\left(\frac{1}{J_i}\hat{\sigma}_i^2\right)^2\right] \\
&\quad + \mathbb{E}\left[c_i^2\right]\mathbb{E}\left[u_i^2\left(\theta - \mathbb{E}\left[\theta_i\right]\right)^2\right] - 2\beta^2 V\mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2 c_i^2\left(c_i - c_i^2\right)\right] \\
&= \mathbb{E}\left[u_i^2\left(\theta - \mathbb{E}\left[\theta_i\right]\right)^2\right] + o(1), \qquad (7)
\end{aligned}$$

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

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

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

which is derived from (7) and the dominated convergence theorem, then we have

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

By (7) and Slutsky's theorem,

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

Therefore,

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

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

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

□

Proof of Theorem 3.3. Firstly, for simplicity we abbreviate notations and denote $\hat{\beta}_{c,\mathrm{HE}}$ as $\hat{\beta}_c$, $\hat{\beta}_{\mathrm{HE}}$ as $\hat{\beta}$, $\hat{\theta}_{i,c,\mathrm{HE}}$ as $\hat{\theta}_{i,c}$, $\hat{\theta}_{i,\mathrm{HE}}$ as $\hat{\theta}_i$ and $\mathrm{Var}\left(\theta_i\right)$ 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

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

and then show for the denominator part

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

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:

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

the first order derivative of which is denoted as

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

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 $\left|\tilde{V} - V\right| \leq \left|\hat{V} - V\right|$,

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

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

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

Then we only need to show

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

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,

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

where the last limit follows from the independence and finite moments assumptions, as well as the fact from Lemma B.1 that $\mathbb{E}\left[\left|\bar{X}_i\right|^{2k_1}\right]\to\mathbb{E}\left[\left|\theta_i\right|^{2k_1}\right]$. 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

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

Therefore,

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

Then the numerator part is shown.

Lastly, we show for the denominator part

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

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

$$g_n\left(\hat{V}\right)-g_n(V)=\left(\hat{V}-V\right)g_n'\left(\tilde{V}\right).$$

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

$$g_n'\left(\tilde{V}\right)=O_p(1).$$

Let

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

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

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

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

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

Because

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

therefore it is proved. □

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

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

Then by Lemma C.1, in order to show

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

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

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

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

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

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

$$f_n\left(\hat{V}\right)-f_n\left(V\right)=\left(\hat{V}-V\right)f_n'\left(\tilde{V}\right).$$

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

$$f_n'\left(\tilde{V}\right)=O_p(1).$$

Let

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

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

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

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

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

Because

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

therefore it is proved.

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

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

□

References

Abdulkadiroğlu, 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., 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. Sølvsten (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 $\mathrm{Var}(\theta_i)$ as V.

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

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

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

$$\begin{aligned} \mathrm{Var}\left(\hat{\sigma}_i^2 \mid \sigma_i^2, J_i\right) &= \frac{1}{J_i} \mathbb{E}\left[\epsilon_{i,j}^4 \mid \sigma_i^2, J_i\right] - \frac{\sigma_i^2 (J_i - 3)}{J_i (J_i - 1)} \\ &\le \frac{1}{J_i} K \sigma_i^4 - \frac{\sigma_i^2 (J_i - 3)}{J_i (J_i - 1)} && \text{(Assumption 3.1.3)} \\ &\le \frac{1}{J_i} K \sigma_i^4, && \text{(Assumption 3.1.1)} \qquad (8) \end{aligned}$$

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 $k_1 \in \{1, 2\}$, $k_2 \ge 1$, we have

   (a) $\bar{\epsilon}_i = O_p(n^{-1/4})$, *and* $\mathbb{E}\left[|\bar{\epsilon}_i|^{k_1}\right] \to 0$.

   (b) $\bar{X}_i = \theta_i + O_p(n^{-1/4})$, *and* $\mathbb{E}\left[\left|\bar{X}_i\right|^{k_1}\right] \to \mathbb{E}\left[|\theta_i|^{k_1}\right]$.

   (c) $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$, *and* $\mathbb{E}\left[|\hat{\sigma}_i^2 - \sigma_i^2|^{k_1}\right] \to 0$.

   (d) $\mathbb{E}\left[\frac{1}{J_i} \hat{\sigma}_i^2\right] \to 0$, *and* $\mathbb{E}\left[\left(\frac{1}{J_i} \hat{\sigma}_i^2\right)^2\right] \to 0$.

   (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* $\mathbb{E}\left[c_i^{k_2}\right] = 1 + O(n^{-1/2})$.

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

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

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

   (a) $\bar{\theta} = \mathbb{E}\left[\theta_i\right] + O_p(n^{-1/2})$.

   (b) $\bar{X} = \mathbb{E}\left[\theta_i\right] + O_p(n^{-1/2})$.

   (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,

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

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

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

[2] For 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 (8),

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

Due to (3), we have $\hat{\sigma}_i^2=\sigma_i^2+O_p(n^{-1/4})$.

Since by (8),

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

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

(d) We have

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

$$\mathbb{E}\left[\left(\frac{1}{J_i}\hat{\sigma}_i^2\right)^2\right]\leq\mathbb{E}\left[\frac{1}{J_i^2}\left(\frac{K\sigma_i^4}{J_i}+\sigma_i^4\right)\right]\to 0.\quad\text{(By (8))}$$

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

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

where $\mu_k, k = 1, ..., 8$ 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 (4), we have that

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

(f) By the mean value theorem,

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

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

$$\begin{aligned}\left|\sqrt{n}\mathbb{E}\left[1-c_i^{k_2}\right]\right| &\leq \frac{k_2}{V}\sqrt{n}\mathbb{E}\left[\frac{\sigma_i^2}{J_i}\right] \\ &\quad + \frac{k_2\left(k_2+1\right)}{2V^2}\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\left(\sigma_i^4+\frac{K}{J_i}\sigma_i^4\right)\right] && \text{(By (8))} \\ &\leq \frac{k_2}{V}\mathbb{E}\left[\sigma_i^2\right]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i}\right] \\ &\quad + \frac{k_2\left(k_2+1\right)}{2V^2}\mathbb{E}\left[\sigma_i^4\right]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^2}\right] \\ &\quad + \frac{k_2\left(k_2+1\right)K}{2V^2}\mathbb{E}\left[\sigma_i^4\right]\sqrt{n}\mathbb{E}\left[\frac{1}{J_i^3}\right] && \text{(By Assumption 3.1.1)} \\ &\to \frac{k_2}{V}\mathbb{E}\left[\sigma_i^2\right]\kappa+0+0. && \text{(By (3) and (4))}\end{aligned}$$

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),

$$\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}c_i^{k_1}\left(\theta_i-k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2}-\mathbb{E}\left[c_i^{k_1}\left(\theta_i-k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2}\right]\right|>s\right)$$
$$\leq\frac{\mathbb{E}\left[c_i^{2k_1}\left(\theta_i-k_3\mathbb{E}\left[\theta_i\right]\right)^{2k_2}\right]}{s^2}\leq\frac{\mathbb{E}\left[\left(\theta_i-k_3\mathbb{E}\left[\theta_i\right]\right)^{2k_2}\right]}{s^2}$$

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

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

Meanwhile, by independence from Assumption 3.1.3,

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

by the mean value theorem ($k_4 = 0$) and (8),

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

where the last step follows from independence of Assumption 3.1.1, (3), (4) and 1e,

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

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

and also by Assumption 3.1.3,

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

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

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

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

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

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

   (c) By Chebyshev's inequality and (8), for any $s > 0$,

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

   and also

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

   Thus

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

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

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

Thus

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

(b) By Chebyshev's inequality,

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

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

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

Then the properties of the difference are

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

For c_i^2, by the mean value theorem,

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

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

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

Since we have independence of $\sigma_i^2 \perp\!\!\!\perp J_i$ from Assumption 3.1.1, (3) (4), and that

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

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

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

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

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

And the proof is complete. □

Recall that we define in the beginning of this section $V \coloneqq \mathrm{Var}\left(\theta_i\right)$ and in Section 3.3

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

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,

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

□

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 $\text{Var}\left(\theta_i\right)$ as V.

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

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

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

From Lemma B.3,

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

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

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

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

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

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

$$- 2\sqrt{n}\left(\bar{\theta} - \mathbb{E}[\theta_i]\right)\bar{\epsilon} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\epsilon}_i^2$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\theta_i - \mathbb{E}[\theta_i])^2 + o_p(1).$$

Therefore, the numerator is

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

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

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

□

C Lemmas and Proofs for Inference

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

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

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

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

□

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

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

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

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

$$\mathbb{E}\left[c_i^{k_1}\theta_i^{k_3}\right] = \mathbb{E}\left[c_i^{k_1}\right]\mathbb{E}\left[\theta_i^{k_3}\right] \to \mathbb{E}\left[\theta_i^{k_3}\right]. \qquad \text{(By Lemma B.1 1f)}$$

For $k_4 = 4$, by independence and (8),

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

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

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

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

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

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

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

Therefore for $1 \le k \le 4$, $k_2 \ge 0$,

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

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

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

The above also applies to $k = 0$.

Therefore,

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

□

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

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

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,

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

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

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

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

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

□

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

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

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

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

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^{n} c_i^{k_1} \bar{X}_i^{k_2} u_i - \mathbb{E}\left[c_i^{k_1} \bar{X}_i^{k_2} u_i\right]\right| > s\right) \leq \frac{\mathbb{E}\left[\bar{X}_i^{2k_2} u_i^2\right]}{ns^2} \to 0. \tag{9}$$

For $0 \leq k_3 \leq 3$,

$$\mathbb{E}\left[c_i^{k_1} \theta_i^{k_3} u_i\right] = \mathbb{E}\left[c_i^{k_1}\right] \mathbb{E}\left[\theta_i^{k_3} u_i\right] \to \mathbb{E}\left[\theta_i^{k_3} u_i\right]. \qquad \text{(By Lemma B.1 1f)}$$

For $k_4 = 4$, by independence and (8),

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

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

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

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

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

Therefore,

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

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

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

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

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

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

For $0 \leq k_3 \leq 3$,

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

For $k_4 = 4$, by independence and (8),

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

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

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

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

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

Therefore,

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

□

D Lemmas and Proofs for Correlated J_i and σ_i^2

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

We next define the shrinkage weight

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

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

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

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

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

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

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

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

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

Assumption D.2.

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

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

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

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* $\mathbb{E}\left[|\bar{\epsilon}_i|^{k_1}\right] \to 0$.

   (b) $\bar{X}_i = \theta_i + O_p(n^{-1/4})$, *and* $\mathbb{E}\left[\left|\bar{X}_i\right|^{k_1}\right] \to \mathbb{E}\left[|\theta_i|^{k_1}\right]$.

   (c) $\hat{\sigma}_i^2 = \sigma_i^2 + O_p(n^{-1/4})$, *and* $\mathbb{E}\left[|\hat{\sigma}_i^2 - \sigma_i^2|^{k_1}\right] \to 0$.

   (d) $\mathbb{E}\left[\frac{1}{J_i}\hat{\sigma}_i^2\right] \to 0$, *and* $\mathbb{E}\left[\left(\frac{1}{J_i}\hat{\sigma}_i^2\right)^2\right] \to 0$.

   (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) $c_{i,\mathrm{HO}}^{k_2} = 1 + O_p(n^{-1/2})$, *and* $\mathbb{E}\left[c_{i,\mathrm{HO}}^{k_2}\right] = 1 + O(n^{-1/2})$.

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

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

$$\frac{1}{n}\sum_{i=1}^{n} c_i^{k_1}\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2} = \mathbb{E}\left[\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2}\right] + O_p(n^{-1/2}),$$

$$\frac{1}{n}\sum_{i=1}^{n} c_i^{k_1}\left(\theta_i - \mathbb{E}\left[\theta_i\right]\right)^{k_4}\bar{\epsilon}_i = o_p(n^{-1/2}),$$

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

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

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

   (a) $\bar{\theta} = \mathbb{E}\left[\theta_i\right] + O_p(n^{-1/2})$.

   (b) $\bar{X} = \mathbb{E}\left[\theta_i\right] + O_p(n^{-1/2})$.

   (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,$c_{i,HO}$:

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

[3] For 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 $c_{i,\mathrm{HO}}$.

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,

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

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

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

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

   (b) This follows from 1a.

   (c) By Markov's inequality and (8),

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

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

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

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

(d) We have

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

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

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

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

where $\mu_k, k = 1, ..., 8$ 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

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

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,

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

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

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

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,

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

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),

$$\begin{aligned}&\Pr\left(\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n} c_i^{k_1}\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2} - \mathbb{E}\left[c_i^{k_1}\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{k_2}\right]\right| > s\right)\\ &\le \frac{\mathbb{E}\left[c_i^{2k_1}\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{2k_2}\right]}{s^2} \le \frac{\mathbb{E}\left[\left(\theta_i - k_3\mathbb{E}\left[\theta_i\right]\right)^{2k_2}\right]}{s^2}\end{aligned}$$

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

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

Meanwhile,

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

by the mean value theorem ($k_4 = 0$) and (8),

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

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

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

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

and also by Assumption 3.1.3,

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

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

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

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

(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 (8),

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

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 $u_i \perp\!\!\!\perp \epsilon_{i,j} \mid \theta_i$, by Chebyshev's inequality,

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

Thus

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

(b) By Chebyshev's inequality,

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

Thus

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

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

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

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

Then the properties of the difference are

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

For c_i^2, by the mean value theorem,

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

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

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

Since we have Assumption D.2 and that

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

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

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

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

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

And the proof for c_i is complete.

For $c_{i,\mathrm{HO}}$, 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,

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

□

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,

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

□

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

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

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*

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

Proof of Proposition D.2. For the shrinkage weight

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

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

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

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

where the denominator

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

and the numerator

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

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

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

Now, focusing on the bias term, by Chebyshev's inequality and (8), for any $s > 0$,

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

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

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

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

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

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

which proved 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

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

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

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

Then by Chebyshev's inequality,

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

which proves the second result of the theorem. □