

Inference with Many Weak Instruments and Heterogeneity

Luther Yap*

September 26, 2024

LATEST VERSION [HERE](#).

Abstract

This paper considers inference in a linear instrumental variable regression model with many potentially weak instruments and heterogeneous treatment effects. I first show that existing test procedures, including those that are robust to only either weak instruments or heterogeneous treatment effects, can be arbitrarily oversized in this setup. Then, I propose a valid inference procedure based on a score statistic and a leave-three-out variance estimator. To establish this procedure’s validity, this paper proves that the score statistic is asymptotically normal and the variance estimator is consistent. The power of the score test is also close to a power envelope in an empirical application.

1 Introduction

Many empirical studies in economics involve instrumental variables (IV) models with many instruments. A prominent example is the judge design: several studies use randomly assigned judges or case workers as instruments to study the effects of foster care (Doyle, 2007), incarceration (Aizer and Doyle, 2015), detention (Dobbie et al., 2018), disability benefits (Autor et al., 2019), and prosecution (Agan et al., 2023), among others. Another example of many IV is a single instrument fully interacted with discrete covariates (e.g., Angrist and Krueger (1991)). Moreover, existing empirical surveys (e.g., Andrews et al. (2019); Lee et al. (2022)) tend to understate how common this setting is: they document that about half the regressions in their sample use a just-identified IV setting, but some papers that construct a “leniency measure” for judges are (correctly) classified as being just-identified with a scalar IV, even though the many instruments complications arise.

Despite the pervasiveness of this setting, there does not yet exist an inference procedure that is robust to both heterogeneous treatment effects and weak instruments, which is a gap this paper aims to fill. Weak instruments here refers to a setting where no consistent estimator for the object of interest exists; and heterogeneous treatment effects refers to a setting where different subsets of the many instruments may estimate different local average treatment effects. It is well-understood

*Email: lyap@princeton.edu.

the standard two-stage least squares (TSLS) procedure for IV is biased, inconsistent, and yields invalid inference in the many-instrument environment (e.g., [Bekker \(1994\)](#); [Donald and Newey \(2001\)](#)). While the jackknife IV estimator (JIVE) (e.g., [Phillips and Hale \(1977\)](#); [Angrist et al. \(1999\)](#); [Chao et al. \(2012\)](#)) solves the estimation problem because it is unbiased, it does not solve the over-rejection problem of TSLS due to weak IV. There are several recent proposals ([Crudu et al., 2021](#); [Mikusheva and Sun, 2022](#); [Matsushita and Otsu, 2022](#)) that make JIVE robust to weak IV, but they assume constant treatment effects. A separate literature ([Kolesár, 2013](#); [Evdokimov and Kolesár, 2018](#)) has adapted JIVE to be robust to heterogeneous treatment effects, but it is not robust to weak IV. While it is clear that weak IV can lead to substantial distortions in inference (e.g., [Staiger and Stock \(1997\)](#)), it is less obvious if procedures developed under constant treatment effects that are robust to weak IV are still valid with heterogeneous treatment effects.

In this paper, I first show that neglecting either heterogeneity or weak instruments can result in substantial distortions in inference. Section 2 presents a simple simulation that has both weak instruments and heterogeneous treatment effects. For a nominal 5% test, using the procedure from [Mikusheva and Sun \(2022\)](#) (henceforth MS22) that is robust to weak instruments but not heterogeneity can result in 100% rejection under the null, because their test statistic is not centered correctly when there is heterogeneity. This result is attributed to how their test is a joint test of both the parameter value and the null of no heterogeneity. Similarly, the procedure from [Evdokimov and Kolesár \(2018\)](#) (henceforth EK18), which is robust to heterogeneity but not weak instruments, can result in 100% rejection when there are weak instruments.

Given the stark simulation results, Section 3 proposes a procedure for valid inference. Following the many instruments literature, the JIVE estimand is the object of interest — this estimand can be interpreted as a weighted average of treatment effects when there is heterogeneity (e.g., EK18). Using weak identification asymptotics, I show that the Lagrange Multiplier (LM) (i.e., score) statistic, earlier proposed by [Matsushita and Otsu \(2022\)](#) under constant treatment effects, is mean zero and asymptotically normal even with treatment effect heterogeneity. In fact, I prove a stronger normality result that a set of jackknife statistics that includes the LM is jointly normal, which is the first technical challenge of this paper. This normality implies that, as long as the variance of LM is consistently estimable, a t statistic can be calculated and critical values from the standard normal distribution are valid for inference. I prove that a leave-three-out (L3O) variance estimator, motivated by the procedure in [Anatolyev and Sølvesten \(2023\)](#) for the OLS problem with many covariates, is consistent. Consistency of L3O is the second technical challenge of this paper. Even in an environment where the reduced-form coefficients are not consistently estimable, this variance estimator consistently estimates the LM variance. Due to the generality of the setting considered, beyond its robustness to weak IV and heterogeneity, the procedure proposed in this paper is also robust to (i) heteroskedasticity, (ii) potentially few observations per instrument, and (iii) potentially many covariates, so it retains the advantages of existing procedures in the literature.

Section 4 argues that the proposed LM procedure is reasonably powerful. In the environment with a fixed reduced-form covariance matrix, I show theoretically that the one-sided LM test is the

most powerful test for testing the null against any alternative from a well-defined set and that the two-sided LM test is the uniformly most powerful unbiased test, at least in the class of tests that are functions of a natural set of statistics. To investigate the power properties of one-sided tests beyond the set, I construct a power envelope (i.e., the highest achievable power for any feasible test that is valid across the entire null space for a given alternative) using the algorithm in [Elliott et al. \(2015\)](#) (henceforth EMW). Using an adversarial covariance matrix, the power of the LM test can be substantially lower than the power envelope. Calibrating the variance matrix based on the empirical application of [Angrist and Krueger \(1991\)](#), LM is essentially indistinguishable from the EMW power envelope. This numerical evidence suggests that LM is not necessarily uniformly most powerful, but it is nonetheless simple and informative, and the empirical gains from a more powerful procedure are likely small in the given asymptotic problem.

The robustness of the validity results is assessed using simulations in Section 6 that involve small sample sizes, binary treatment (which implies non-normal residuals), and a setting with covariates. The LM procedure with L3O variance has close to nominal coverage under a judge structural model. I also implement my proposed procedure in the [Angrist and Krueger \(1991\)](#) quarter of birth application in Section 5, and show that the confidence interval is wide, but their result is nonetheless significant.

This paper contributes to the following strands of literature. First, this paper contributes to a growing literature on many weak instruments. There is a strand of literature dealing with many instruments (e.g., [Chao and Swanson \(2005\)](#); [Chao et al. \(2012\)](#)) and another separate strand on weak instruments (e.g., [Staiger and Stock \(1997\)](#); [Stock and Yogo \(2005\)](#); [Lee et al. \(2023\)](#)). While recent work accommodates both simultaneously (e.g., [Crudu et al. \(2021\)](#); [Mikusheva and Sun \(2022\)](#); [Matsushita and Otsu \(2022\)](#); [Yap \(2023\)](#); [Lim et al. \(2024\)](#)), its focus has been on the linear IV model with constant treatment effects. This paper augments their setup by allowing for heterogeneity in treatment effects.

Second, this paper contributes to the literature on heterogeneous treatment effects (e.g., [Kolesár \(2013\)](#); [Evdokimov and Kolesár \(2018\)](#); [Blandhol et al. \(2022\)](#)). These papers provide conditions where the coefficient of interest can be consistently estimated, and exploit that consistency to conduct inference. In this paper, I operate in the (more general) weak IV environment so the object of interest may not be consistently estimated, but I still have sufficient conditions for inference. One paper that allows weak IV and heterogeneity is contemporaneous work in [Boot and Nibbering \(2024\)](#), who study a single discrete instrument interacted and saturated with many covariates. Their setup is a special case of the environment considered in this paper and the many weak instruments literature, so it is unclear if their procedure generalizes to many instruments without covariates (e.g., judges). Additionally, I characterize power properties of the score statistic.

Third, this paper contributes to a literature on inference when coefficients cannot be consistently estimated. The difficulty in having such a general robust inference procedure lies in consistent variance estimation when the number of coefficients is large. Recent literature that has made substantial progress in a different context. In doing inference in OLS with many covariates, [Cattaneo](#)

et al. (2018) and Anatolyev and Solvsten (2023) proposed consistent variance estimators that are robust to heteroskedasticity, which involve inverting a large (n by n , where n is the sample size) matrix (similar to Hartley et al. (1969)) and a L3O approach respectively. Boot and Nibbering (2024) adapt the Cattaneo et al. (2018) variance estimator for inference. In contrast, this paper adapts the approach from Anatolyev and Solvsten (2023) that does not require an inversion of an n by n matrix, and whose L3O implementation is fast when using matrix operations.

In the rest of this paper, Section 2 explains how existing procedures are invalid using a simple simulation. Section 3 proposes a valid inference procedure. Section 4 discusses the power properties of the score statistic; Section 5 discusses implementation issues; Section 6 presents further simulation results; Section 7 concludes. Implementation code can be found at: <https://github.com/lutheryap/mwivhet>.

2 Challenges in Conventional Practice

This section explains the challenges faced in conventional practice by considering a simple linear potential outcomes model without covariates that exhibits weak instruments and heterogeneity in treatment effects. A simulation from the model shows how weak instruments and heterogeneity can lead to substantial distortions in inference. In particular, for a nominal 5% test, MS22 (which is not valid under heterogeneity) can result in 100% rejection in a data-generating process (DGP) with heterogeneity; EK18 (which is not valid under weak instruments) can result in 100% rejection in a DGP with weak instruments. A common empirical practice of constructing a leave-one-out instrument and then applying inference methods for the instrument as if it is not constructed also has high rejection rates. In contrast, the method proposed in this paper (in the following section) has a rejection rate that is close to the nominal rate.

2.1 Setting for Simple Example

Consider a linear potential outcomes model with a K -dimensional instrument Z that is a vector of indicators for judges with $K \rightarrow \infty$, each with $c = 5$ cases, a continuous endogenous variable X , and outcome Y :

$$X_i(z) = z' \pi + v_i, \text{ and } Y_i(x) = x(\beta + \xi_i) + \varepsilon_i. \quad (1)$$

In this model, the individual treatment effect is $\beta_i = \beta + \xi_i$. We can interpret v_i as the noise associated with the first-stage regression, ε_i as the noise in the intercept of the outcome equation, and ξ_i as the individual-level treatment effect heterogeneity. This model is a special case of the model in Section 3, which presents an inference procedure that is valid for a more general model that also accommodates (potentially many) covariates, heteroskedasticity, and distributions of residuals that are not normal. The joint distribution of $(v_i, \xi_i, \varepsilon_i)$ conditional on judge k is:

$$\begin{pmatrix} \varepsilon_i \\ \xi_i \\ v_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\xi} & \sigma_{\varepsilon v} \\ \cdot & \sigma_{\xi\xi} & \sigma_{\xi v k} \\ \cdot & \cdot & \sigma_{vv} \end{pmatrix} \right). \quad (2)$$

Due to the judge design, $X_i = \pi_{k(i)} + v_i$, where $k(i)$ is the judge that observation i is assigned to. The strength of the instrument is $S = \frac{1}{\sqrt{K}} \sum_k (c-1) \pi_k^2$. The π_k 's are constructed as such: with $s = \sqrt{S/\sqrt{K}/(c-1)}$, set $\pi_0 = 0$ for the base judge, $\pi_k = -s$ for half the judges and $\pi_k = s$ for the other half. The heterogeneity covariances $\sigma_{\xi v k}$ are constructed so that $\sum_k \pi_k = 0$, $\sum_k \sigma_{\xi v k} = 0$, and $\sum_k \pi_k \sigma_{\xi v k} = 0$. With C_H characterizing the heterogeneity in the model, and $h = \sqrt{C_H/\sqrt{K}/(c-1)}$, set $\sigma_{\xi v 0} = 0$; among judges with $\pi_k = s$, half of them have $\sigma_{\xi v k} = h$ and the other half $\sigma_{\xi v k} = -h$. The same construction of $\sigma_{\xi v k}$ applies for judges with $\pi_k = -s$. In simulations with constant treatment effects, I set $\xi_i = 0$ for all i , so $\sigma_{\xi\xi} = \sigma_{\varepsilon\xi} = \sigma_{\xi v} = 0$.

Remark 1. In this model, $\sigma_{\xi v k}$ characterizes the extent of treatment effect heterogeneity. The observed outcome in a model with constant treatment effects is $Y_i(X_i) = X_i\beta + \varepsilon_i$, with $E[\varepsilon_i] = 0$. When $\sigma_{\xi v k} = 0$, regardless of the values of $\sigma_{\varepsilon\xi}$, $\sigma_{\xi\xi}$, the observed outcome of Equation (1) can be written as $Y_i(X_i) = X_i\beta + \varepsilon_i$ where $E[\varepsilon_i] = E[X_i\xi_i + \varepsilon_i] = E[X_i E[\xi_i | X_i]] = 0$, which resembles the constant treatment effect case.

The object of interest is the average treatment effect β . For every design, I run 1000 simulations and calculate the frequency that each inference procedure rejects the null of $\beta_0 = 0$. These procedures include the standard TSLS, procedures that are robust to weak instruments, procedures that are robust to heterogeneity, and procedures that use a constructed instrument. The results are presented in Table 1, which I will refer to in the remainder of this section.

From the table, it is evident that, if we ignore the many IV problem and simply run TSLS for an over-identified model, inference is invalid with 100% rejection in many designs, a fact already known in the literature. It is also known that the TSLS estimator is biased and inconsistent. In TSLS, the first stage regresses X on Z to get a predicted $\hat{X} = Z\hat{\pi}$, where $\hat{\pi}$ is the estimated coefficient; the second stage regresses Y on \hat{X} . In this example, the bias of the TSLS estimator is $E[\varepsilon_i Z_i' \hat{\pi}] = (K/n) \sigma_{\varepsilon v}$, where K/n is non-negligible for $c = 5$, and $\sigma_{\varepsilon v}$ is nonzero in general. The problem arises because using i to estimate $\hat{\pi}$ generates a bias when constructing $\hat{X}_i = Z_i' \hat{\pi}$.

A natural solution to the bias in the TSLS estimator is to use the JIVE to estimate β . Instead of using $\hat{X}_i = Z_i' \hat{\pi}$ in the second stage, we instead use $\tilde{X}_i = Z_i' \hat{\pi}_{-i}$, where $\hat{\pi}_{-i}$ is the coefficient from the first-stage regression that leaves out observation i . I will also call this the leave-one-out (L1O) coefficient. With $P = Z(Z'Z)^{-1}Z'$ denoting the projection matrix, $\tilde{X}_i = Z_i' \hat{\pi}_{-i}$ can be written as $\tilde{X}_i = \sum_{j \neq i} P_{ij} X_j$. Then, the JIVE is:

$$\hat{\beta} = \frac{\sum_i Y_i \left(\sum_{j \neq i} P_{ij} X_j \right)}{\sum_i X_i \left(\sum_{j \neq i} P_{ij} X_j \right)}.$$

Table 1: Rejection rates under the null for nominal size 0.05 test for continuous X

	TSLS	EK	MS	MO	JIVEC	ARC	L3O	LMorc
$C_H = S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$	0.061	0.017	1.000	0.079	0.079	0.078	0.042	0.044
$C_H = 2\sqrt{K}, S = 2\sqrt{K}$	0.952	0.022	1.000	0.082	0.087	0.084	0.058	0.055
$C_H = 2\sqrt{K}, S = 2$	1.000	0.009	1.000	0.125	0.076	0.127	0.053	0.050
$C_H = 2\sqrt{K}, S = 0$	1.000	0.006	1.000	0.128	0.061	0.127	0.059	0.052
$C_H = 3, S = 3\sqrt{K}$	0.986	0.033	0.109	0.060	0.062	0.064	0.056	0.047
$C_H = 3, S = 3$	1.000	0.036	0.168	0.065	0.078	0.087	0.055	0.047
$C_H = 3, S = 0$	1.000	0.048	0.184	0.066	0.106	0.088	0.053	0.057
$C_H = 0, S = 2\sqrt{K}$	1.000	0.089	0.049	0.068	0.083	0.080	0.061	0.058
$C_H = 0, S = 2$	1.000	0.207	0.045	0.076	0.243	0.135	0.057	0.045
$C_H = 0, S = 0$	1.000	0.337	0.051	0.062	0.413	0.127	0.045	0.048
$C_H = S = 0, \sigma_{\varepsilon v} = 1$	1.000	1.000	0.044	0.061	1.000	0.157	0.052	0.044

Table Notes: Unless mentioned otherwise, simulations use $K = 500, c = 5, \beta = 0, \sigma_{\varepsilon\varepsilon} = \sigma_{vv} = 1, \sigma_{\varepsilon\xi} = 0, \sigma_{\varepsilon v} = 0.8, \sigma_{\xi\xi} = 1 + h$ for $h^2 < 1$ with 1000 simulations. The table displays rejection rates of various procedures (in columns) for various designs (in rows). $C_H = 0$ uses $\xi_i = 0$ for all i , which uses $\sigma_{\xi\xi} = \sigma_{\varepsilon\xi} = \sigma_{\varepsilon v} = 0$, corresponding to constant treatment effects. TSLS implements the standard two-stage-least-squares procedure for an over-identified IV system. EK implements the procedure in [Evdokimov and Kolesár \(2018\)](#). MS uses the statistic in [Mikusheva and Sun \(2022\)](#) with an oracle variance of their statistic. MO uses the variance estimator proposed in [Matsushita and Otsu \(2022\)](#). JIVEC uses a constructed instrument and runs TSLS for a just-identified IV system. ARC uses the AR procedure for a just-identified system using a constructed instrument. L3O uses the variance estimator proposed in this paper. LMorc is the infeasible theoretical benchmark that uses an LM statistic with an oracle variance.

The JIVE is unbiased under constant treatment effects, and is commonly used in existing procedures for many IV that are robust to either heterogeneity or weak instruments, which I turn to next.

2.2 Issue with Weak Instruments

It is well-known that, in the just-identified IV context with weak instruments, the rejection rate of the standard t-ratio can be up to 100% for a nominal 5% test (e.g., [Dufour \(1997\)](#); [Staiger and Stock \(1997\)](#)). In the many IV context with constant treatment effects, the asymptotic distribution of the t-statistic of the JIVE is the same as the distribution of the t-statistic of the TSLS estimator ([Mikusheva and Sun, 2022](#); [Yap, 2023](#)) — it is a ratio of two normally distributed random variables. Hence, like the just-identified IV context, by using a structural model that has sufficiently weak instruments and high covariance, the simulation can deliver rejection rates of 100%.

EK18 have a procedure that is robust to heterogeneity, but not weak instruments, so even if we use their variance estimator for the t statistic, this problem is not alleviated. This fact is evident in the EK column of Table 1 where the rejection rate increases to 1 as the instrument S gets weaker and there is perfect correlation with $\sigma_{\varepsilon v} = 1$. Hence, ignoring the issue of weak instruments can lead to substantial distortions in inference. In fact, even with strong instruments, there is no guarantee that EK18 achieves the nominal rate, because their variance estimation method requires consistent

estimation of the first-stage coefficients $\hat{\pi}$. A condition for consistent variance estimation is that the number of cases per judge is large, which is not $c = 5$.

2.3 Issue with Heterogeneity

Next, we consider proposals for inference that are developed for contexts with many weak instruments, of which there are several. MS22 (and [Crudu et al. \(2021\)](#)) propose using an Anderson-Rubin (AR) statistic $T_{ee} = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} e_i e_j$, for $e_i := Y_i - X_i \beta_0$ where β_0 is the hypothesized null value. This statistic is motivated by how e_i is the null-imposed residual: if the instrument is orthogonal to the residual, then $E[Z'e] = 0$. Then, T_{ee} is the L1O analog for the quadratic form that tests the moment $E[Z'e] = 0$. Since observations are independent, the critical value for the test is obtained from a mean-zero normal distribution. In this model,

$$E[T_{ee}] = \frac{1}{\sqrt{K}} \sum_k (c-1) \sigma_{\xi vk}^2 = \sqrt{K}(c-1)h^2. \quad (3)$$

Hence, when there are constant treatment effects such that $\sigma_{\xi vk} = 0$ for all k , the statistic is unbiased. However, in the setup with heterogeneity, the test statistic in MS22 can be biased: in fact, when h does not converge to zero, $E[T_{ee}]$ diverges. Further, there does not exist any estimand β_0 such that $E[T_{ee}] = 0$, as shown in Lemma 4 of the appendix. In the simulation, when h does not converge to 0, the bias is large enough to generate 100% rejection.

Another proposal in the literature that is robust to many weak instruments is [Matsushita and Otsu \(2022\)](#) (henceforth MO22) who use the statistic $T_{eX} = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} e_i X_j$. Since $T_{eX} = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} e_i X_j = \frac{1}{\sqrt{K}} \sum_i e_i \tilde{X}_i$, this statistic can be interpreted as the LM (or score) statistic. They propose the following variance estimator $\hat{\Psi}_{MO}$:

$$\hat{\Psi}_{MO} := \sum_i \left(\sum_{j \neq i} P_{ij} X_j \right)^2 e_i^2 + \sum_i \sum_{j \neq i} P_{ij}^2 X_i e_i X_j e_j. \quad (4)$$

While T_{eX} has zero mean, the variance is constructed under constant treatment effects, so the variance estimand differs from the true variance. In particular, it is shown in the appendix that, for this model, $E[\hat{\Psi}_{MO}] \neq \text{Var}(T_{eX})$ in general. When $\beta = \beta_0 = 0$,

$$\begin{aligned} \text{Var}(T_{eX}) = & \sum_k \frac{(c-1)^2}{c} \omega_{\zeta \zeta k} \pi_k^2 + \sum_k \frac{c-1}{c} \left(\omega_{\zeta \zeta k} \sigma_{vv} + (\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v})^2 \right) \\ & + 2 \sum_k \frac{(c-1)^2}{c} (\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v}) \pi_k \sigma_{\xi vk} + \sum_k \frac{(c-1)^2}{c} \sigma_{vv} \sigma_{\xi vk}^2, \text{ and} \end{aligned} \quad (5)$$

$$\begin{aligned}
E \left[\hat{\Psi}_{MO} \right] &= \sum_k \frac{(c-1)^2}{c} \pi_k^2 \sigma_{\xi vk}^2 + \sum_k \frac{(c-1)^2}{c} \pi_k^2 \omega_{\zeta \zeta k} + \sum_k \frac{c-1}{c} \omega_{\zeta \zeta k} \sigma_{vv} + \sum_k \frac{c-1}{c} \sigma_{\xi vk} \sigma_{vv} \\
&\quad + \sum_k \frac{c-1}{c} \left(\pi_k^2 \sigma_{\xi vk}^2 + 2(\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v}) \pi_k \sigma_{\xi vk} + (\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v})^2 \right),
\end{aligned} \tag{6}$$

where $\omega_{\zeta \zeta k} := \pi_k^2 \sigma_{\xi \xi} + 2\pi_k \sigma_{\varepsilon \xi} + \sigma_{\varepsilon \varepsilon} + \sigma_{vv} \sigma_{\xi \xi} + \sigma_{\xi vk}^2$. When $\sigma_{\xi vk} = 0$, we have $E \left[\hat{\Psi}_{MO} \right] = \text{Var} (T_{eX})$. The variance estimator $\hat{\Psi}_{MO}$ is inconsistent in general, so when it is used to construct the t-statistic of T_{eX} , the normalized statistic is not distributed $N(0, 1)$ asymptotically. Consequently, by constructing a DGP where $\hat{\Psi}_{MO}$ underestimates the variance, it is possible to get over-rejection of the MO22 procedure, as in the cases of Table 1 where C_H diverges. As expected, when there is no heterogeneity such that $h = 0$, the rejection rate of MO22 and MS22 are close to the nominal rate. MS22 is closer to the nominal rate than MO22 because I used an oracle variance for MS22 and an estimated variance for MO22.

2.4 Issue with a Constructed Instrument

In light of problems with weak identification and heterogeneity, a possible response is to transform a many instruments environment into a just-identified single-instrument environment. With a single instrument, the [Anderson and Rubin \(1949\)](#) procedure (among others) is robust to both weak identification and heterogeneity. However, this subsection will argue that such an approach is invalid.

Due to how the JIVE is written, there are several empirical papers that treat $\tilde{X}_i = \sum_{j \neq i} P_{ij} X_j$ as the “instrument” so that $\hat{\beta} = \sum_i Y_i \tilde{X}_i / \sum_i X_i \tilde{X}_i$, and proceed with inference as if \tilde{X}_i is not constructed, but is an observed scalar instrument, usually referred to as the leniency measure. While the resulting estimator is numerically identical to JIVE, there are distortions in inference because the variance estimators do not account for the variability in constructing \tilde{X}_i .

If the TSLS t-ratio inference is used as if \tilde{X}_i is the instrument, then the rejection rates are somewhat worse than using EK18 that accounts for the variance accurately, by comparing the JIVEC and EK columns in Table 1.

Even if the weak-instrument robust AR procedure for just-identified instruments were used, there can still be distortion in inference (see the ARC column of Table 1). To see how the distortion arises, the AR t-statistic is $t_{ARC} := \sum_i e_i \tilde{X}_i / \sqrt{\hat{V}}$, where $\hat{V} = \sum_i \tilde{X}_i^2 \hat{e}_i^2 / \left(\sum_i \tilde{X}_i^2 \right)^2$ and $\hat{e}_i = e_i - \tilde{X}_i \left(\sum_i e_i \tilde{X}_i \right) / \left(\sum_i \tilde{X}_i^2 \right)$. Even though t_{ARC} is mean zero and asymptotically normal, the variance estimand is inaccurate, much like MO22. In particular, when $\beta = 0$, the leading term of the variance is:

$$E \left[\sum_i \tilde{X}_i^2 e_i^2 \right] = \sum_k \frac{(c-1)^2}{c} \pi_k^2 \sigma_{\xi v}^2 + \sum_k \frac{c-1}{c} \sigma_{\xi v}^2 \sigma_{vv} + \sum_k \frac{(c-1)^2}{c} \pi_k^2 \omega_{\zeta \zeta k} + \sum_k \frac{c-1}{c} \omega_{\zeta \zeta k} \sigma_{vv}. \tag{7}$$

The estimand is different from the true variance stated in Equation (5), and the estimator can be inconsistent. Hence, using the just-identified AR procedure with a constructed instrument results

in over-rejection. As a preview, the L3O procedure proposed in this paper has rejection rates close to the nominal rate while the other procedures can over-reject.

3 Valid Inference

In light of how existing procedures are invalid in an environment with many weak instruments and heterogeneity as documented in the previous section, this section describes a novel inference procedure and shows that it is valid. I set up a general model, then show that an LM statistic is asymptotically normal and a feasible variance estimator is consistent, which suffices for inference.

3.1 Setting: Model and Asymptotic Distribution

The general setup mimics [Evdokimov and Kolesár \(2018\)](#). With an independently drawn sample of individuals $i = 1, \dots, n$, we observe each individual's scalar outcome Y_i , scalar endogenous variable X_i , instrument Z_i , and covariates W_i , where $\dim(Z_i) = K$. For every instrument value z , there is an associated potential treatment $X_i(z)$, and we observe $X_i = X_i(Z_i)$. Similarly, potential outcomes are denoted $Y_i(x)$, with $Y_i = Y_i(X_i)$. Let $R_i = E[X_i | Z_i, W_i]$ and $R_{Yi} = E[Y_i | Z_i, W_i]$, and these are assumed to be linear. The model, written in the reduced-form and first-stage equations, is:

$$\begin{aligned} Y_i &= R_{Yi} + \zeta_i, \text{ where} & R_{Yi} &:= Z_i' \pi_Y + W_i' \gamma_Y, & E[\zeta_i | Z_i, W_i] &= 0, \text{ and} \\ X_i &= R_i + \eta_i, \text{ where} & R_i &:= Z_i' \pi + W_i' \gamma, & E[\eta_i | Z_i, W_i] &= 0. \end{aligned}$$

The setup implicitly conditions on Z_i, W_i , so R_i, R_{Yi} are nonrandom.¹ This model implies linearity in Z and W , which is not necessarily restrictive when K is large, because any nonlinear function of the instruments can be arbitrarily well-approximated by a spline with a large number of pieces or a high-order polynomial. Moreover, the arguments in this paper could presumably be extended to a linear approximation of nonlinear functions as long as there are regularity conditions to ensure that higher-order terms are asymptotically negligible.

Define $e_i := Y_i - X_i \beta$, where β is some estimand of interest that we want to do inference on, and e_i is an associated linear transformation. Further, let $R_{\Delta i} := R_{Yi} - R_i \beta$ and $\nu_i := \zeta_i - \eta_i \beta$. These definitions imply that $e_i = R_{\Delta i} + \nu_i$ and $R_{\Delta i} = Z_i'(\pi_Y - \pi \beta) + W_i'(\gamma_Y - \gamma \beta)$. Since $E[\nu_i | Z_i, W_i] = 0$ from the model, $E[e_i | Z_i, W_i] = R_{\Delta i}$, which need not be zero. For data matrix A , let $H_A = A(A'A)^{-1}A'$ denote the hat matrix and $M_A = I - H_A$ its corresponding annihilator matrix. With Z, W denoting the corresponding data matrices of the instrument and covariates, let $Q = (Z, W)$, $P = H_Q$, and $M = I - P$. C is used to denote arbitrary constants.

Remark 2. While $E[e_i | Z_i, W_i] = R_{\Delta i}$ need not be zero under heterogeneous treatment effects, $E[e_i | Z_i, W_i] = R_{\Delta i} = 0$ under constant treatment effects. Since $R_{\Delta i} = Z_i'(\pi_Y - \pi \beta) + W_i'(\gamma_Y - \gamma \beta)$

¹If we are interested in a superpopulation where Z is random, then the estimands would be defined as the probability limit of the conditional objects. Then, it suffices to have regularity conditions to ensure that the conditional object converges to the unconditional object.

for all i , constant treatment effects with $E[Y_i - X_i\beta \mid Z_i, W_i] = 0$ also implies $\pi_Y = \pi\beta$ and $\gamma_Y = \gamma\beta$ outside of edge cases (e.g., when Z_i, W_i are always 0). These Δ objects hence capture the impact of having heterogeneous treatment effects in the many instruments model.

The (conditional) object of interest and its corresponding estimator are:

$$\beta_{JIVE} := \frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j}, \text{ and } \hat{\beta}_{JIVE} = \frac{\sum_i \sum_{j \neq i} G_{ij} Y_i X_j}{\sum_i \sum_{j \neq i} G_{ij} X_i X_j},$$

where G is an $n \times n$ matrix that can take several forms. As the leading cases, if there are no covariates, $G = H_Z$ is simply the projection matrix, and when there are covariates, I use the unbiased JIVE “UJIVE” (Kolesár, 2013) with $G = (I - \text{diag}(H_Q))^{-1} H_Q - (I - \text{diag}(H_W))^{-1} H_W$.² In an environment with a binary instrument and many covariates interacted with the instrument, the saturated estimand “SIVE” (Boot and Nibbering, 2024) uses $G = P_{BN} - M_Q D_{BN} M_Q$, where $P_{BN} = M_W Z (Z' M_W Z)^{-1} Z' M_W$ and D_{BN} is defined as a diagonal matrix with elements such that $P_{BN, ii} = [M_Q D_{BN} M_Q]_{ii}$. With constant treatment effects, the estimand is the same for all the estimators: $R_{Yi} = R_i\beta$ so $\beta_{JIVE} = \beta$. Depending on the application, the estimand is usually interpretable as some weighted average of treatment effects (Evdokimov and Kolesár, 2018).³ The focus of this paper is on inference, so I will not discuss the estimand in detail. The results for valid inference in the paper are established for any G that satisfies properties that will be formally stated in the theorem. For simplicity, other properties (e.g., power) are analyzed without covariates.

This paper restricts its attention to the following statistics:

$$(T_{ee}, T_{eX}, T_{XX})' := \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (e_i e_j, e_i X_j, X_i X_j)'$$

These T objects are observed because the e_i objects can be calculated by using the null-imposed β . It suffices to focus on (T_{ee}, T_{eX}, T_{XX}) for inference as they correspond to a linear transformation of the leave-one-out analog of a maximal invariant — details are in Section 4.1. T_{ee} is the (unnormalized) AR statistic used by MS22 for inference, and T_{eX} is the LM (score) statistic used by MO22. T_{XX} corresponds to a first-stage F statistic.

The following assumption states sufficient conditions for joint asymptotic normality.

Assumption 1. (a) There exists $C < \infty$ such that $E[\eta_i^4] + E[\nu_i^4] \leq C$ for all i .

(b) $E[\nu_i^2]$ and $E[\eta_i^2]$ are bounded away from 0 and $|\text{corr}(\nu_i, \eta_i)|$ is bounded away from 1.

(c) As $K \rightarrow \infty$, there exists $\underline{c} > 0$ such that $\left(\frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \right)^2 \geq \underline{c}$.

²Even if Z includes full interaction of a discrete instrument (say quarter of birth) and W , there is still value in partialling out W . The main difference is that, if for a given covariate group, all observations have the same instrument value, then UJIVE will not incorporate those observations at all. In contrast, merely using Z will still incorporate these observations.

³In the judge example without covariates above, we have $G = P$ and $\pi_{Yk} = \beta_k \pi_k$ where β_k is the local average treatment effect (LATE) between judge k and the base judge, so $\beta_{JIVE} = \frac{\sum_k \pi_{Yk} \pi_k}{\sum_k \pi_k^2} = \frac{\sum_k \pi_k^2 \beta_k}{\sum_k \pi_k^2}$ is a weighted average of LATE's.

- (d) $\frac{1}{K^2} \sum_i \left(\left(\sum_{j \neq i} G_{ij} R_j \right)^4 + \left(\sum_{j \neq i} G_{ij} R_{\Delta j} \right)^4 + \left(\sum_{j \neq i} G_{ji} R_j \right)^4 + \left(\sum_{j \neq i} G_{ji} R_{\Delta j} \right)^4 \right) \rightarrow 0.$
- (e) $\| \frac{1}{K} G_L G'_L \|_F + \| \frac{1}{K} G_U G'_U \|_F \rightarrow 0$, where G_L is a lower-triangular matrix with elements $G_{L,ij} = G_{ij} 1\{i > j\}$ and G_U is an upper-triangular matrix with elements $G_{U,ij} = G_{ij} 1\{i < j\}$.

Assumption 1 states high-level conditions that mimic EK18 so that a central limit theorem (CLT) can be applied. These conditions hence accommodate the G that EK18 consider with covariates. Having bounded moments in (a) is standard. Conditions (b) and (c) are sufficient to ensure that the variance is non-zero asymptotically. In particular, (b) rules out perfect correlation: in the simulation, $\text{corr}(\eta_i, \nu_i) = -1$ is the pathological case that makes the variance zero, but $\text{corr}(\eta_i, \nu_i) = 1$ still allows non-zero variance. Condition (c) imposes a many instruments environment with $K \rightarrow \infty$ and ensures that $\sum_i \sum_{j \neq i} G_{ij}^2$ is of order at least K . Conditions (d) and (e) ensure that the weights placed on the individual stochastic terms are not too large.

The conditions on G are satisfied when $G = P$ is a projection matrix. For (c), any rank K projection matrix satisfies $\sum_i \sum_j P_{ij}^2 = K$. Due to Lemma B3 of [Chao et al. \(2012\)](#), under weak IV asymptotics where $P_{ii} \leq C < 1$, Assumption 1(e) is satisfied, as $\|G_L G'_L\|_F \leq C\sqrt{K}$. Notably, the conditions do not require $P_{ii} \rightarrow 0$ so the π, π_Y coefficients need not be consistently estimated.

Theorem 1. *Under Assumption 1, let $S_K = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_i R_j$. Then,*

$$\hat{\beta}_{JIVE} - \beta_{JIVE} = \frac{\frac{1}{\sqrt{K}} \left(\sum_i \sum_{j \neq i} G_{ij} (R_{\Delta i} \eta_j + \nu_i R_j + \nu_i \eta_j) \right)}{S_K + \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j)} = \frac{T_{eX}}{T_{XX}},$$

and for some variance matrix V , as $K \rightarrow \infty$ and $n \rightarrow \infty$,

$$\begin{pmatrix} T_{ee} - \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_{\Delta i} R_{\Delta j} \\ T_{eX} \\ T_{XX} - \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_i R_j \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, V \right). \quad (8)$$

Theorem 1 states a numerical equivalence on the difference between $\hat{\beta}_{JIVE}$ and β_{JIVE} . Under strong identification, $\hat{\beta}_{JIVE}$ converges in probability to β_{JIVE} , but we no longer have such convergence under weak instruments (i.e., with [Staiger and Stock \(1997\)](#) asymptotics). S_K is the concentration parameter corresponding to the instrument strength. If the instruments are strong, then $S_K \rightarrow \infty$, so $\hat{\beta}_{JIVE} - \beta_{JIVE} \xrightarrow{d} 0$. With weak IV, S_K converges to some constant $C < \infty$.

The asymptotic distribution follows from establishing a quadratic CLT that may be of independent interest: it is proven by rewriting sums as a martingale difference array, then applying the martingale CLT. Using that result, with $K \rightarrow \infty$, the distribution holds regardless of instrument strength. While there are existing quadratic CLT available, they do not fit the context exactly. [Chao et al. \(2012\)](#) Lemma A2 requires G to be symmetric, which works without covariates as it is just a projection matrix, but G for UJIVE is not symmetric. EK18 Lemma D2 is established for scalar random variables, so I extend it to random vectors.

This theorem implies that, under weak identification, comparing the JIVE t statistic with the standard normal distribution leads to invalid inference even in large samples. The theorem also states that the asymptotic distribution is a ratio of two normals, which is identical to the distribution of the just-identified TSLS estimator. While [Yap \(2023\)](#) and MS22 have observed this result in part with many weak instruments, their results are restricted to the case with constant treatment effects. Here, I show that the distribution holds even with heterogeneous treatment effects. Theorem 1 also states that T_{eX} is mean zero and asymptotically normal in this general environment. Hence, if we have access to the oracle variance of T_{eX} , we can simply use the statistic $T_{eX}/\sqrt{\text{Var}(T_{eX})}$ for testing because it has a standard normal distribution under the null. Obtaining a consistent estimator is an issue addressed in the next subsection.

3.2 Variance Estimation

To test the null that $H_0 : \beta = \beta_0$, we can calculate T_{eX} using the null-imposed β_0 and an estimator for the variance of $\sqrt{K}T_{eX}$, \hat{V}_{LM} , defined later in this section. Then, reject if $KT_{eX}^2/\hat{V}_{LM} \geq \Phi(1 - \alpha/2)^2$ for a size α test where $\Phi(\cdot)$ is the standard normal CDF. This procedure is valid when T_{eX} is asymptotically normal with mean zero as we have established in the previous section, and when \hat{V}_{LM} is consistent.

Before stating the variance estimator, I first decompose the variance expression in the equation below, which follows from substituting $e_i = R_{\Delta i} + \nu_i$ and $X_i = R_i + \eta_i$ into the variance. It is shown in the appendix that, for $V_{LM} := \text{Var}\left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j\right)$,

$$\begin{aligned} V_{LM} = & \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\nu_i^2] G_{ij} G_{ik} R_j R_k + \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E[\eta_i \nu_i] E[\eta_j \nu_j] \\ & + 2 \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\nu_i \eta_i] G_{ij} G_{ki} R_j R_{\Delta k} + \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\eta_i^2] G_{ji} G_{ki} R_{\Delta j} R_{\Delta k}. \end{aligned} \quad (9)$$

As a corollary, if $G = P$ is a projection matrix and $M = I - P$, then

$$\begin{aligned} \text{Var}\left(\sum_i \sum_{j \neq i} P_{ij} e_i X_j\right) = & \sum_i E[\nu_i^2] M_{ii}^2 R_i^2 + \sum_i \sum_{j \neq i} P_{ij}^2 (E[\nu_i^2] E[\eta_j^2] + E[\eta_i \nu_i] E[\eta_j \nu_j]) \\ & + 2 \sum_i E[\nu_i \eta_i] M_{ii}^2 R_i R_{\Delta i} + \sum_i E[\eta_i^2] M_{ii}^2 R_{\Delta i}^2. \end{aligned} \quad (10)$$

With constant treatment effects, only the first line appears in the variance. The expression for $\text{Var}\left(\sum_i \sum_{j \neq i} P_{ij} e_i X_j\right)$ matches the expression in EK18 Theorem 5.3, but their variance estimator cannot be used directly as they required consistent estimation of reduced-form coefficients. By adapting the leave-three-out (L3O) approach of [Anatolyev and S¸olvsten \(2023\)](#) (henceforth AS23), an unbiased and consistent variance estimator can be obtained. Let $\tau := (\pi', \gamma')'$ and $\tau_{\Delta} := ((\pi_Y - \pi\beta)', (\gamma_Y - \gamma\beta'))'$ denote the coefficients on Q when running the regression of X and e

respectively. In the following, for notational convenience, let $M = M_Q$. The variance estimator is:

$$\hat{V}_{LM} := \sum_i (A_{1i} + 2A_{2i} + A_{3i}) - \sum_i \sum_{j \neq i} G_{ji}^2 A_{4ij} - \sum_i \sum_{j \neq i} G_{ij} G_{ji} A_{5ij}, \quad (11)$$

with

$$\begin{aligned} A_{1i} &:= \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ik} X_k e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk}), \\ A_{2i} &:= \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ki} e_k e_i (X_i - Q'_i \hat{\tau}_{-ijk}), \\ A_{3i} &:= \sum_{j \neq i} \sum_{k \neq i} G_{ji} e_j G_{ki} e_k X_i (X_i - Q'_i \hat{\tau}_{-ijk}), \\ A_{4ij} &:= X_i \sum_{k \neq j} \check{M}_{ik, -ij} X_k e_j (e_j - Q'_j \hat{\tau}_{\Delta, -ijk}), \\ A_{5ij} &:= e_i \sum_{k \neq j} \check{M}_{ik, -ij} X_k e_j (X_j - Q'_j \hat{\tau}_{-ijk}), \end{aligned}$$

where

$$\begin{aligned} \hat{\tau}_{-ijk} &:= \left(\sum_{l \neq i, j, k} Q_l Q'_l \right)^{-1} \sum_{l \neq i, j, k} Q_l X_l, \\ \hat{\tau}_{\Delta, -ijk} &:= \left(\sum_{l \neq i, j, k} Q_l Q'_l \right)^{-1} \sum_{l \neq i, j, k} Q_l e_l, \\ D_{ij} &:= M_{ii} M_{jj} - M_{ij}^2, \text{ and} \\ \check{M}_{ik, -ij} &:= \frac{M_{jj} M_{ik} - M_{ij} M_{jk}}{D_{ij}} = -Q'_i \left(\sum_{l \neq i, j} Q_l Q'_l \right)^{-1} Q_k. \end{aligned}$$

Following AS23, I make an assumption to ensure that the L3O estimator is well-defined.⁴

Assumption 2. (a) $\sum_{l \neq i, j, k} Q_l Q'_l$ is invertible for every $i, j, k \in \{1, \dots, n\}$.

(b) $\max_{i \neq j \neq k \neq i} D_{ijk}^{-1} = O_P(1)$, where $D_{ijk} := M_{ii} D_{jk} - (M_{jj} M_{ik}^2 + M_{kk} M_{ij}^2 - 2M_{jk} M_{ij} M_{ik})$.

Assumption 2(a) corresponds to AS23 Assumption 1 and Assumption 2(b) corresponds to AS23 Assumption 4. For consistent variance estimation, we additionally require regularity conditions that are stated in Assumption 3 of the appendix. These conditions are satisfied when G is a projection matrix. With these conditions, Theorem 2 below claims that the variance estimator is consistent.

Theorem 2. Under Assumptions 1-2, and Assumption 3 in the appendix, $E[\hat{V}_{LM}] = V_{LM}$ and the variance estimator is consistent, i.e., $\hat{V}_{LM}/V_{LM} \xrightarrow{p} 1$.

⁴If these conditions are not satisfied, then we can follow the modification in AS23 so that the variance estimator is conservative.

With many instruments and potentially many covariates, the main difficulty is that the reduced-form coefficients $\pi, \pi_Y, \gamma, \gamma_Y$ are not consistently estimable. The usual approach to constructing variance estimators calculates residuals by using the estimated coefficients. When these estimated coefficients are inconsistent, standard arguments for consistency of the variance estimator can no longer be applied. To be precise, applying Chebyshev's inequality for any $\epsilon > 0$ yields:

$$\Pr \left(\left| \frac{\hat{V}_{LM} - V_{LM}}{V_{LM}} \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \frac{\text{Var}(\hat{V}_{LM})}{V_{LM}^2} + \frac{1}{\epsilon^2} \frac{(E[\hat{V}_{LM}] - V_{LM})^2}{V_{LM}^2}. \quad (12)$$

Without an unbiased estimator and when reduced-form coefficients cannot be consistently estimated, the second term in (12) is not necessarily asymptotically negligible. To overcome this problem, I use an unbiased variance estimator so that the second term is exactly zero. Then, it suffices to show that the variance of individual components of the variance are asymptotically small compared to V_{LM}^2 , so that the first term in (12) is $o(1)$ by applying the Cauchy-Schwarz inequality.

To obtain an unbiased estimator, I use estimators for the reduced-form coefficients $\pi, \pi_Y, \gamma, \gamma_Y$ that are unbiased and independent of objects that they are multiplied with, which helps to construct an unbiased variance estimator. The leave-three-out (L3O) approach provides this unbiasedness for linear regressions: when leaving three observations out in the inner-most sum, the estimated coefficient $\hat{\tau}_{-ijk}$ is independent of i, j, k and is unbiased for τ . Then, when taking the expectation through a product of random variables of i, j, k and $\hat{\tau}_{-ijk}$, τ can be used in place of the $\hat{\tau}_{-ijk}$ component. This helps to recover the triple sums in the V_{LM} expression of Equation (9). An analogous argument applies to other components of V in (8) so they can be consistently estimated.

4 Power Properties

This section characterizes power properties of the valid LM procedure. To do so, I first argue that we can restrict our attention to three statistics that are jointly normal. Since the reduced-form covariance can be consistently estimated, the remainder of the section then focuses on the 3-variable normal distribution with a known covariance matrix. With this asymptotic distribution, I qualify some theoretical optimality results on one-sided and two-sided LM tests. Namely, the one-sided LM test is the most powerful test against alternatives within a subset and the two-sided LM test is the uniformly most powerful unbiased test within the interior of the parameter space. Beyond these qualifications, I derive the power envelope more generally using the algorithm from [Elliott et al. \(2015\)](#). While LM may not achieve the power envelope in general, I find that, when using a covariance matrix calibrated from the [Angrist and Krueger \(1991\)](#) empirical application, LM numerically achieves the power envelope for local alternatives.

4.1 Sufficient Statistics and Maximal Invariant

As is standard in the literature, I consider the canonical model without covariates where the reduced-form errors are normal and homoskedastic (e.g., [Andrews et al. \(2006\)](#); [Moreira \(2009a\)](#); [Mikusheva and Sun \(2022\)](#)). In this environment, I derive a maximal invariant and its associated distribution for the reduced-form model without covariates. Suppose that (η, ζ) are jointly normal with known variance. To be precise,

$$\begin{pmatrix} \zeta_i \\ \eta_i \end{pmatrix} \sim N(0, \Omega) = N\left(0, \begin{bmatrix} \omega_{\zeta\zeta} & \omega_{\zeta\eta} \\ \omega_{\zeta\eta} & \omega_{\eta\eta} \end{bmatrix}\right). \quad (13)$$

Define:

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} := \begin{pmatrix} (Z'Z)^{-1/2} Z'Y \\ (Z'Z)^{-1/2} Z'X \end{pmatrix}.$$

I restrict attention to tests that are invariant to rotations of Z , i.e., transformations of the form $Z \rightarrow ZF'$ where F is a $K \times K$ orthogonal matrix. In particular, an invariant test $\phi(s_1, s_2)$ is one for which $\phi(Fs_1, Fs_2) = \phi(s_1, s_2)$ for all $K \times K$ orthogonal matrices F . In the judges case with constant treatment effects, it is reasonable to expect the test to yield the same result regardless of which judge we choose to normalize with. If we focus on invariant tests, then the maximal invariant contains all relevant information from the data for inference.

Lemma 1. $(s'_1, s'_2)'$ are sufficient statistics for $(\pi'_Y, \pi')'$. Further, for transformations of the form $Z \rightarrow ZF'$ where F is a $K \times K$ orthogonal matrix, $(s'_1 s_1, s'_1 s_2, s'_2 s_2)$ is a maximal invariant, and

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \sim N\left(\begin{pmatrix} (Z'Z)^{1/2} \pi_Y \\ (Z'Z)^{1/2} \pi \end{pmatrix}, \Omega \otimes I_K\right).$$

The derivation for Lemma 1 mimics [Moreira \(2009a\)](#) Proposition 4.1. After demeaning appropriately, the maximal invariant $(s'_1 s_1, s'_1 s_2, s'_2 s_2)$ is jointly normal, as claimed in the proposition below.

Proposition 1. With Equation (13), $K \rightarrow \infty$ and $\frac{1}{\sqrt{K}} (\pi'_Y Z' Z \pi_Y, \pi' Z' Z \pi_Y, \pi' Z' Z \pi) \rightarrow (C_{YY}, C_Y, C_S)$,

$$\frac{1}{\sqrt{K}} \begin{pmatrix} s'_1 s_1 - K\omega_{\zeta\zeta} - C_{YY} \\ s'_1 s_2 - K\omega_{\zeta\eta} - C_Y \\ s'_2 s_2 - K\omega_{\eta\eta} - C_S \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma\right) \quad (14)$$

for some variance matrix Σ . If $C_{YY}, C_Y, C_S < \infty$,

$$\Sigma = \begin{pmatrix} 2\omega_{\zeta\zeta}^2 & 2\omega_{\zeta\eta}\omega_{\zeta\zeta} & 2\omega_{\zeta\eta}^2 \\ 2\omega_{\zeta\eta}\omega_{\zeta\zeta} & \omega_{\zeta\zeta}\omega_{\eta\eta} + \omega_{\zeta\eta}^2 & 2\omega_{\zeta\eta}\omega_{\eta\eta} \\ 2\omega_{\zeta\eta}^2 & 2\omega_{\zeta\eta}\omega_{\eta\eta} & 2\omega_{\eta\eta}^2 \end{pmatrix}.$$

The proof of Proposition 1 relies on $K \rightarrow \infty$ because objects like $s'_1 s_1$ can be written as a sum of

K objects. With an appropriate representation to obtain independence, a law of large numbers and a central limit theorem can be applied to yield normality. Compared to MS22, Proposition 1 does not require constant treatment effects and characterizes the distribution without orthogonalizing the sufficient statistics. Nonetheless, the form of the covariance matrix is similar to MS22.

Considering the leave-one-out (L1O) analog of the maximal invariant is attractive in this context because it removes the need to subtract the variance objects on the left-hand side of Equation (14). Without covariates such that $G = P$, I define $(T_{YY}, T_{YX}, T_{XX}) := \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} (Y_i Y_j, Y_i X_j, X_i X_j)$. This (T_{YY}, T_{YX}, T_{XX}) is the L1O analog of the maximal invariant $(s'_1 s_1, s'_1 s_2, s'_2 s_2)$.⁵ This L1O analog also relates to JIVE directly because $\hat{\beta}_{JIVE} = T_{YX}/T_{XX}$. As a corollary of Theorem 1, since (T_{YY}, T_{YX}, T_{XX}) is a linear transformation of (T_{ee}, T_{eX}, T_{XX}) that is jointly normal, (T_{YY}, T_{YX}, T_{XX}) is also jointly normal.⁶ In light of how (T_{YY}, T_{YX}, T_{XX}) is the L1O analog of the maximal invariant and it is related directly to JIVE, I restrict our attention to tests that are functions of (T_{YY}, T_{YX}, T_{XX}) .

4.2 Discussion of Asymptotic Problem

The asymptotic problem involving (T_{YY}, T_{YX}, T_{XX}) is:

$$\begin{pmatrix} T_{YY} \\ T_{YX} \\ T_{XX} \end{pmatrix} \sim N(\mu, \Sigma), \mu = \begin{pmatrix} \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_{Yj} \\ \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_{Xj} \\ \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{Xi} R_{Xj} \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \cdot & \sigma_{22} & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{pmatrix}. \quad (15)$$

There are several natural restrictions in the μ vector, which is assumed to be finite. Since P is a projection matrix, $\sum_i \sum_{j \neq i} P_{ij} R_{Xi} R_{Xj} = \sum_i R_{Xi} (\sum_j P_{ij} R_{Xj} - P_{ii} R_{Xi}) = \sum_i M_{ii} R_{Xi}^2$. Since the annihilator matrix M has positive entries on its diagonal, we obtain $\mu_3 \geq 0$ and a similar argument yields $\mu_1 \geq 0$. Further, with $\mu_2 = \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_{Xj} = \sum_i M_{ii} R_{Yi} R_{Xi}$, the Cauchy-Schwarz inequality implies $\mu_2^2 \leq \mu_1 \mu_3$. Notably, constant treatment effects implies $\mu_2^2 = \mu_1 \mu_3$, which is a special case of the environment here. These properties do not contradict the joint normality: even though $\mu_3 \geq 0$, T_{XX} can still be negative when using the L1O statistic.

Beyond the necessary restrictions that $\mu_1, \mu_3 \geq 0$ and $\mu_2^2 \leq \mu_1 \mu_3$, there is also a question of whether Σ places further restrictions on μ , which can give more information about $\beta_{JIVE} = \mu_2/\mu_3$. While Σ is uninformative when we have normal homoskedastic reduced-form errors, it is less obvious if there exists any structural model where this result still holds when β features in Σ . With more structure, there can be more restrictions on μ , but if there is no structural model where Σ is uninformative, then any necessary restriction should be accounted for in the asymptotic problem.

The main results in this subsection establish that there exists a structural model where Σ is uninformative about μ , and $\mu_1, \mu_3 \geq 0$ and $\mu_2^2 \leq \mu_1 \mu_3$ are the *only* restrictions on μ . In particular, the model of Section 2 with weak identification and weak heterogeneity where $\sqrt{K} s^2 \rightarrow C_S < \infty$

⁵To see this, $s'_1 s_2 = Y' Z (Z' Z)^{-1} Z' X = Y' P X = \sum_i \sum_j P_{ij} Y_i X_j$.

⁶To see that (T_{YY}, T_{YX}, T_{XX}) is a linear transformation, use the fact that $e = Y + X\beta$. Then, $(T_{YY}, T_{YX}, T_{XX}) := \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} ((e_i + X_i \beta)(e_j + X_j \beta), (e_i + X_i \beta) X_j, X_i X_j) = (T_{ee} + 2T_{eX}\beta + T_{XX}\beta^2, T_{eX} - T_{XX}\beta, T_{XX})$.

and $\sqrt{K}h^2 \rightarrow C_H < \infty$ is one such example. Consequently, it suffices to focus on the 3-variable normal model with the non-negativity and Cauchy-Schwarz restrictions on μ .

The following lemma expresses the reduced-form objects in terms of structural parameters, and the proposition formally states the result.

Lemma 2. *Consider the model of Section 2. If $\sqrt{K}s^2 \rightarrow C_S < \infty$ and $\sqrt{K}h^2 \rightarrow C_H < \infty$, then*

$$\sigma_{11} = \frac{4}{\sigma_{33}} \left(\sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right)^2 + o(1), \quad \sigma_{12} = 2 \frac{\sigma_{23}}{\sigma_{33}} \left(\sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right) + o(1), \quad \sigma_{13} = \frac{\sigma_{23}^2}{\sigma_{33}} + o(1),$$

$$\sigma_{22} = \frac{c-1}{c} \left(\sigma_{vv} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\varepsilon v}\beta) + (\sigma_{vv}\beta + \sigma_{\varepsilon v})^2 \right) + o(1),$$

$$\sigma_{33} = 2 \frac{c-1}{c} \sigma_{vv}^2 + o(1), \quad \sigma_{23} = 2 \frac{c-1}{c} \sigma_{vv} (\sigma_{vv}\beta + \sigma_{\varepsilon v}) + o(1), \text{ and}$$

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \sqrt{K}(c-1)(s^2\beta^2 + h^2) \\ \sqrt{K}(c-1)s^2\beta \\ \sqrt{K}(c-1)s^2 \end{pmatrix} = (c-1) \begin{pmatrix} C_S\beta^2 + C_H \\ C_S\beta \\ C_S \end{pmatrix}.$$

Proposition 2. *In the model of Section 2 with $\sqrt{K}s^2 \rightarrow C_S < \infty$ and $\sqrt{K}h^2 \rightarrow C_H < \infty$, for any $\sigma_{22}, \sigma_{23}, \sigma_{33}$ such that $\sigma_{22}, \sigma_{33} > 0$, $\sigma_{23}^2 \leq \sigma_{22}\sigma_{33}$ and μ such that $\mu_1, \mu_3 > 0$, $\mu_2^2 \leq \mu_1\mu_3$, the following values of structural parameters:*

$$\begin{aligned} C_S &= \mu_3 / (c-1), \quad \beta = \mu_2 / \mu_3, \quad h = \sqrt{\frac{1}{\sqrt{K}} \frac{1}{c-1} \left(\mu_1 - \frac{\mu_2^2}{\mu_3} \right)}, \\ \Sigma_{SF} &= \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\xi} & \sigma_{\varepsilon v} \\ . & \sigma_{\xi\xi} & \sigma_{\xi v} \\ . & . & \sigma_{vv} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_{vv}} \frac{c}{c-1} \left(\sigma_{22} - \frac{\sigma_{23}^2}{\sigma_{33}} \right) + \frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}} & 0 & \sigma_{\varepsilon v} \\ . & \frac{h}{\sigma_{vv}} & \pm h \\ . & . & \sigma_{vv} \end{pmatrix}, \\ \sigma_{vv} &= \sqrt{\frac{\sigma_{33}c}{2(c-1)}}, \text{ and } \sigma_{\varepsilon v} = \frac{1}{\sigma_{vv}} \left(\frac{\sigma_{23}c}{2(c-1)} - \sigma_{vv}^2\beta \right), \end{aligned}$$

satisfy the equations in Lemma 2, and $\det(\Sigma_{SF})/h \rightarrow C_{det} \geq 0$.

Due to Proposition 2, since the principal submatrices of Σ_{SF} are positive semidefinite asymptotically, Σ_{SF} is a symmetric positive semidefinite matrix. The proposition thus implies that when the σ 's and μ satisfy the conditions, there exists structural parameters that can generate the given μ and Σ asymptotically. Hence, there are no further restrictions on μ from the observed Σ in the simple normal model. The remainder of this section considers how powerful LM is when there are restrictions on the μ vector.

4.3 Analytic Results

Using the asymptotic problem of Equation (15), testing $H_0 : \mu_2/\mu_3 = \beta^*$ is identical to testing $H_0 : \mu_2 - \beta^*\mu_3 = 0$. Since β^* is fixed, and I consider alternatives of the form: $H_A : \mu_2 - \beta^*\mu_3 = h_A$. The LM statistic corresponds to $T_{YX} - \beta^*T_{XX}$, so it can be used to test the null directly. I focus on the most common case of $\beta^* = 0$, and it is analogous to extend the argument for $\beta^* \neq 0$. Having $\beta^* = 0$ simplifies the argument because it suffices to focus on testing the null of $\mu_2 = 0$. Further, $(T_{YY}, T_{YX}, T_{XX}) = (T_{ee}, T_{eX}, T_{XX})$. Let μ^A denote the mean under the alternative and μ^H under the null.

The one-sided and two-sided LM tests are defined in the following manner. With a size α test, the one-sided LM test against the alternative that $\mu_2 > 0$ rejects when $T_{eX}/\sqrt{\text{Var}(T_{eX})} > \Phi(1-\alpha)$. When testing against the alternative that $\mu_2 < 0$, it rejects when $T_{eX}/\sqrt{\text{Var}(T_{eX})} < \Phi(\alpha)$. The two-sided LM test against the alternative that $\mu_2 \neq 0$ rejects when $T_{eX}^2/\text{Var}(T_{eX}) > \Phi(1-\alpha/2)^2$.

The one-sided test is the most powerful test for testing against a particular subset of alternatives \mathcal{S} defined by:

$$\mathcal{S} := \left\{ (\mu_1^A, \mu_2^A, \mu_3^A) : \mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A \geq 0, \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A \geq 0 \right\}.$$

While \mathcal{S} may not be empirically interpretable, this set is constructed so that standard [Lehmann and Romano \(2005\)](#) arguments can be applied to conclude that the one-sided LM test is the most powerful test. For a given $(\mu_1^A, \mu_2^A, \mu_3^A)$ in the alternative space, LM (which just uses the second element) is justified as being most powerful because it is identical to the Neyman-Pearson test when testing against a point null μ^H with $\mu_1^H = \mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A$, $\mu_2^H = 0$ and $\mu_3^H = \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A$. The inequalities in \mathcal{S} are imposed so that $\mu_1^H, \mu_3^H \geq 0$, ensuring that μ^H is in the null space. If these constraints hold, then LM is the most powerful, as the next proposition states. In contrast, if the inequalities fail in the alternative space, then $(\mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A, 0, \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A)$ is not in the null space, and the [Lehmann and Romano \(2005\)](#) argument cannot be applied.

Proposition 3. *The one-sided LM test is the most powerful test for testing any alternative hypothesis $(\mu_1^A, \mu_2^A, \mu_3^A) \in \mathcal{S}$ in the asymptotic problem of Equation (15).*

The proposition makes no statement about alternative hypotheses that are not in \mathcal{S} . In general, LM is not necessarily the most powerful test, as the numerical results in Section 4.4 show. A more powerful test can be constructed when μ_2^A is large and covariance σ_{23}, σ_{12} are large.

Proposition 4. *Consider a restriction of the μ space to the interior i.e., $\mu_1, \mu_3 > 0$ and $\mu_2^2 < \mu_1\mu_3$. Then, the two-sided LM test is the uniformly most powerful unbiased test in the asymptotic problem of Equation (15).*

With some qualifications, the argument for LM being the uniformly most powerful unbiased test from [Moreira \(2009b\)](#) still applies. The argument for optimality applies standard results from [Lehmann and Romano \(2005\)](#) on the exponential family: since the restricted parameter space is

convex, the asymptotic distribution is still in the exponential family. The restriction to the interior is required to apply the theorem, because the [Lehmann and Romano \(2005\)](#) environment requires the existence of alternative values above and below the null value.⁷ Imposing $\mu_1 > 0$ rules out constant treatment effects under the null that $\beta^* = 0$. Further, the proposition claims optimality within the class of unbiased tests, and makes no statement about tests that are biased (i.e., where the power at some point in the alternative space can be lower than the size).

4.4 Numerical Results

While the previous subsection established some optimality properties, there is no reason to be interested only in the set \mathcal{S} for one-sided tests. We may also be interested in weighted average power for two-sided tests rather than the class of unbiased tests. Hence, we turn to numerical analyses to find a power envelope.

The power envelope is achieved by a test that is valid across the entire composite null space, and is most powerful for testing against a particular point in the alternative space. To obtain this test, I implement the algorithm from [Elliott et al. \(2015\)](#) (henceforth EMW) where all weight on the alternative are placed on a single point while being valid across a composite null. Then, testing against every point in the alternative space requires a different critical value. For the numerical exercises in this subsection, I use a Σ matrix of the form:

$$\Sigma = \begin{pmatrix} 2 & 2\rho & 2\rho^2 \\ \cdot & 1 + \rho^2 & 2\rho \\ \cdot & \cdot & 2 \end{pmatrix}, \quad (16)$$

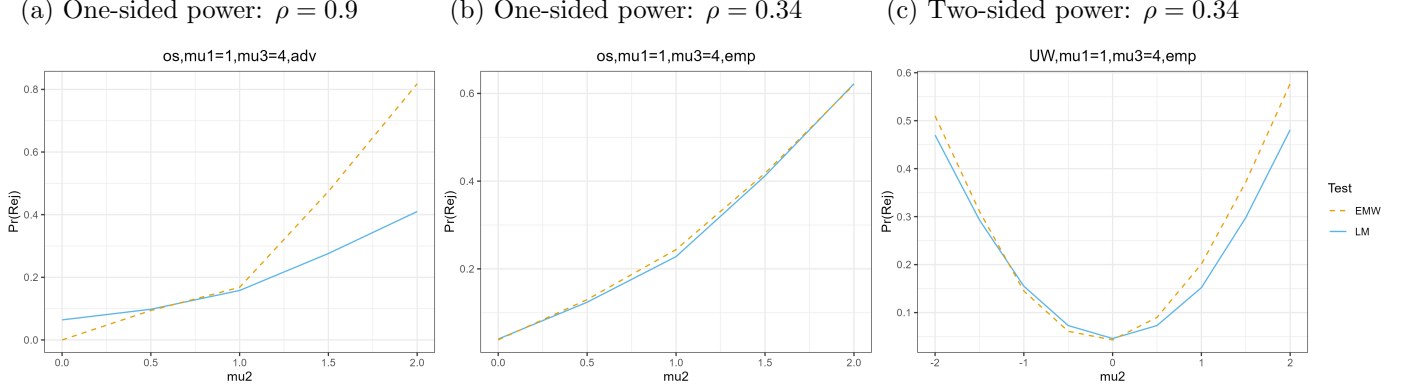
which corresponds to the Σ matrix in Proposition 1 with $\omega_{\zeta\zeta} = \omega_{\eta\eta} = 1, \omega_{\zeta\eta} = \rho$.

The first set of power curves in Figure 1a uses a one-sided LM test and a null grid of $\mu_1 \in [0, 5]$ and $\mu_3 \in [0, 5]$, with a large covariance at $\rho = 0.9$. Every point on the μ_2 axis describes the μ that generated the data, and the null uses $\mu_2^H = 0$. When data is generated from the null, since LM and EMW are valid tests, their rejection rate is at most 0.05. EMW has exact size when testing a weighted average of values in the null space and is valid across the entire space, so when data is generated from one particular point in the null, EMW can be conservative. Consistent with Proposition 3, when μ_2 is small enough, LM achieves the power envelope, but as μ_2 gets larger, the gap widens substantially. This phenomenon occurs because EMW still uses the same null grid, but now it no longer needs to have correct size for testing against the point $(\mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A, 0, \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A)$, as that point is no longer in the null space.

The second set of power curves in Figure 1b uses the same null grid. Σ is calibrated by using the Σ matrix calculated from the [Angrist and Krueger \(1991\)](#) application, so after appropriate normalizations, $\rho = 0.34$. With such a low covariance, LM is basically indistinguishable from the EMW bound. Hence, even though there are gains to be made theoretically, in the empirical

⁷Technically, it suffices to have $\mu_1, \mu_3 > 0$ and $\mu_2^2 \leq \mu_1\mu_3$ when using the null that $\mu_2 = 0$.

Figure 1: Power Curves comparing LM and EMW



Notes: The figures display the rejection rate across 500 independent draws from $X^* \sim N(\mu, \Sigma)$ at each point on the μ_2 axis, with $\mu_1 = 1, \mu_3 = 4$ for a 5% test. The composite null uses a grid of $\mu_1 \in [0, 5], \mu_3 \in [0, 5]$ in 0.5 increments, and assumes the variance is known. In panel (a), $\rho = 0.9$. The LM curve uses the one-sided LM test where we reject if $X_2^*/\sqrt{\sigma_{22}} > \Phi(0.95)$. The EMW curve implements the test in [Elliott et al. \(2015\)](#), where all weight in the alternative space are placed on the given μ^A . Panel (b) has the same design as panel (a), except that $\rho = 0.34$, calibrated from [Angrist and Krueger \(1991\)](#). In panel (c), the two-sided LM tests rejects when $X_2^{*2}/\sigma_{22} > \Phi(0.975)^2$. The alternative grid for EMW places equal weight on alternatives $(\mu_1^A, \mu_2^A, \mu_3^A) \in [0, 5] \times [-2, 2] \times [0, 5]$ in increments of 0.5 (excluding $\mu_2 = 0$) subject to the inequality constraints.

application considered, the gains are small. There are more power curves in [Appendix A.2](#) with different values of μ_1, μ_3 that generate the data.

Instead of considering a point alternative, we may be more interested in testing against a composite alternative. [Figure 1c](#) presents one such possibility by allowing EMW to place equal weight on several points within the alternative space. The resulting test is the nearly optimal test for a weighted average of values the null space against the uniformly weighted average of alternative values. Hence, there is no guarantee that its power is necessarily higher than the LM test at every point in the alternative space. While there are weighted-average power curves in [Appendix A.2](#) that substantially outperform LM like in [Figure 1b](#), they are compatible with [Proposition 4](#). EMW as constructed is a biased test as there are points in the alternative space that are not a part of the grid where LM outperforms EMW. Overall, [Figure 1c](#) suggests that, when using the empirical covariance, LM does not perform substantially worse than EMW.

Remark 3. With the characterized asymptotic distribution, there are several other tests that are valid. (1) We can implement a Bonferroni-type correction that constructs a 99% confidence set for both μ_1 and μ_3 , then a 97% test for LM so that there is a chance for rejection whenever the non-negativity or Cauchy-Schwarz inequality fails. (2) VtF from [Yap \(2023\)](#) can also be implemented, because the asymptotic distribution does not rely on homogeneous treatment effects. There is evidence that it can lead to shorter confidence intervals from [Lee et al. \(2023\)](#). (3) With a given structural model, the EMW grid can also be constructed using a grid on structural parameters.

5 Implementation

5.1 Inversion for Confidence Sets

To invert the LM test to obtain a confidence set, use $e_i = Y_i - X_i\beta_0$ and expand the A expressions in Equation (11) so that we can write everything in terms of X and Y . The two-sided test rejects: $\left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j\right)^2 / \hat{V}_{LM} \leq q = \Phi(1 - \alpha/2)^2$. Let $P_{YX} := \sum_i \sum_{j \neq i} G_{ij} Y_i X_j$. Then, $\sum_i \sum_{j \neq i} G_{ij} e_i X_j = P_{YX} - P_{XX}\beta_0$, so squaring it results in a term that is quadratic in β_0^2 . Similarly, the denominator is quadratic in β_0^2 . Hence, we have triple sums in terms of X, Y that can be calculated in a manner similar to L3O. Then, after calculating it once, we can do an appropriate inversion to get the confidence set.

With $\hat{V}_{LM} = C_0 + C_1\beta_0 + C_2\beta_0^2$, where C_0, C_1, C_2 are coefficients worked out in the appendix, the analysis for the shape of the confidence intervals is similar to the AR procedure for just-identified IV (e.g., Lee et al. (2022)).

Lemma 3. *The test does not reject when $(P_{XX}^2 - qC_2)\beta_0^2 - (2P_{YX}P_{XX} + qC_1)\beta_0 + (P_{YX}^2 - qC_0) \leq 0$. Let*

$$D := (2P_{YX}P_{XX} + qC_1)^2 - 4(P_{XX}^2 - qC_2)(P_{YX}^2 - qC_0).$$

If $D \geq 0$ and $P_{XX}^2 - qC_2 \geq 0$, then the upper and lower bounds of confidence set are:

$$\frac{(2P_{YX}P_{XX} + qC_1) \pm \sqrt{D}}{2(P_{XX}^2 - qC_2)}.$$

If $D < 0$ and $P_{XX}^2 - qC_2 < 0$, then the confidence set is empty. Otherwise, the confidence set is unbounded.

Due to $+qC_1, -qC_2$ in the expression of the upper and lower bounds, the confidence set is not necessarily centered around $\hat{\beta}_{JIVE} = P_{YX}/P_{XX}$.

5.2 Empirical Application

Angrist and Krueger (1991) were interested in the effect of education (X) on wages (Y). They have quarter of birth (QOB) instruments that are fully interacted with covariates: state and year of birth indicators, thereby generating many instruments with this saturation. Implementing the procedure with full interaction (with 2033 instruments) and the full sample (with 329,509 observations), the UJIVE is 0.1027 (vs 0.0928 in their Table VII(2)). The confidence interval is $[0.022, 0.210]$, so the result remains statistically significant, albeit wider than their implied confidence interval.⁸

⁸With a reported standard error of 0.0928 for their TSLS regression, the implied confidence interval is $[0.075, 0.011]$.

Table 2: Rejection Rates under the null for nominal size 0.05 test for Continuous X with $K = 40$

	TSLS	EK	MS	MO	JIVEC	ARC	L3O	LMorc
$C_H = S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$	0.072	0.022	0.525	0.061	0.074	0.068	0.039	0.055
$C_H = 2\sqrt{K}, S = 2\sqrt{K}$	0.238	0.034	0.388	0.066	0.074	0.077	0.055	0.062
$C_H = 2\sqrt{K}, S = 2$	0.547	0.033	0.475	0.111	0.096	0.133	0.077	0.053
$C_H = 2\sqrt{K}, S = 0$	0.651	0.013	0.511	0.094	0.088	0.102	0.068	0.054
$C_H = 3, S = 3\sqrt{K}$	0.213	0.025	0.109	0.057	0.057	0.063	0.055	0.046
$C_H = 3, S = 3$	0.658	0.032	0.129	0.051	0.074	0.063	0.064	0.055
$C_H = 3, S = 0$	0.849	0.049	0.127	0.078	0.109	0.103	0.087	0.057
$C_H = 0, S = 2\sqrt{K}$	0.853	0.105	0.049	0.070	0.068	0.098	0.085	0.056
$C_H = 0, S = 2$	0.999	0.152	0.048	0.062	0.201	0.132	0.098	0.037
$C_H = 0, S = 0$	1.000	0.342	0.052	0.067	0.439	0.143	0.080	0.049
$C_H = S = 0, \sigma_{\varepsilon v} = 1$	1.000	1.000	0.045	0.062	1.000	0.179	0.082	0.045

Note: Designs are identical to Table 1, but $K = 40$ here.

6 Simulations

The general model in Section 3 can be justified by several structural models. Section 2 presented one such structural model with a continuous X and displayed simulation results in Table 1 with $K = 500$. This section discusses the small-sample performance of the various inference procedures in several different structural models.

6.1 Continuous Treatment

This subsection uses the continuous X model of Section 2 and presents results for $K = 40$ in Table 2. The L3O coverage is better than the existing procedures in the literature, albeit worse in smaller samples (when comparing the results with Table 1). MS retains its high rejection rates with strong heterogeneity and EK its high rejection rates with weak instruments. The procedures that use the LM statistic are MO, ARC, L3O and LMorc; they differ only in their variance estimation. Hence, while ARC and MO over-reject, the extent of over-rejection is smaller than MS and EK in the adversarial cases.

6.2 Binary Treatment

This subsection presents a structural model with a binary X . Data is generated from a judge model with $J = K + 1$ judges, each with $c = 5$ cases, and cases are indexed by i . The structural model is:

$$Y_i(x) = x(\beta + \xi_i) + \varepsilon_i, \text{ and}$$

$$X_i(z) = I \{ z' \pi - v_i \geq 0 \},$$

Table 3: Rejection Rates under the null for nominal size 0.05 test for binary X

	TSLS	EK	MS	MO	JIVEC	ARC	L3O	LMorc
$C_H = S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$	0.046	0.049	0.059	0.045	0.045	0.045	0.049	0.054
$C_H = 2\sqrt{K}, S = 2\sqrt{K}$	0.097	0.047	0.177	0.038	0.038	0.041	0.051	0.052
$C_H = 2\sqrt{K}, S = 2$	0.727	0.059	1.000	0.140	0.051	0.143	0.058	0.051
$C_H = 2\sqrt{K}, S = 0$	0.891	0.037	1.000	0.237	0.067	0.247	0.059	0.045
$C_H = 3, S = 3\sqrt{K}$	0.092	0.060	0.051	0.056	0.057	0.056	0.055	0.047
$C_H = 3, S = 3$	0.996	0.089	0.888	0.074	0.086	0.096	0.055	0.048
$C_H = 3, S = 0$	1.000	0.124	0.999	0.128	0.289	0.181	0.068	0.052
$C_H = 0, S = 2\sqrt{K}$	0.408	0.058	0.055	0.043	0.046	0.046	0.045	0.041
$C_H = 0, S = 2$	1.000	0.212	0.052	0.076	0.188	0.108	0.078	0.057
$C_H = 0, S = 0$	1.000	0.654	0.046	0.057	0.750	0.149	0.069	0.039
$C_H = S = 0, \sigma_{\varepsilon\varepsilon} = 0$	1.000	1.000	0.053	0.069	1.000	0.173	0.076	0.053

Note: Unless stated otherwise, designs use $K = 100, c = 5, \beta = 0, p = 7/8, \sigma_{\varepsilon\varepsilon} = 0.1, \sigma_{\varepsilon v} = 0.5$ with 1000 simulations. The procedures in the columns are detailed in the remarks of Table 1.

where $I\{\cdot\}$ is a function that takes value 1 if the argument is true and -1 if false. Our unobservables are generated as follows. Draw $v_i \sim U[-1, 1]$, then generate residuals from:

$$\varepsilon_i \mid v_i \sim \begin{cases} N(\sigma_{\varepsilon v}, \sigma_{\varepsilon\varepsilon}) & \text{if } v_i \geq 0 \\ N(-\sigma_{\varepsilon v}, \sigma_{\varepsilon\varepsilon}) & \text{if } v_i < 0 \end{cases},$$

$$\xi_i \mid v_i \geq 0 = \begin{cases} \sigma_{\xi vk} & w.p. \quad p \\ -\sigma_{\xi vk} & w.p. \quad 1-p \end{cases}, \text{ and } \xi_i \mid v_i < 0 = \begin{cases} \sigma_{\xi vk} & w.p. \quad 1-p \\ -\sigma_{\xi vk} & w.p. \quad p \end{cases}.$$

The process for determining s, h and $\pi_k \in \{0, -s, s\}, \sigma_{\xi vk} \in \{0, -h, h\}$ are identical to Section 2, as s controls the strength of the instrument, h the extent of heterogeneity, and β is the object of interest. Then, the problem's variances and covariances are determined by $(p, \sigma_{\varepsilon v}, \sigma_{\varepsilon\varepsilon})$. The JIVE estimand is shown to be β in this appendix. A simulation is run with $K = 100$, so the sample size is considerably smaller than the normal experiment in Section 2. The procedures that I compare are identical to those mentioned in Section 2.

Results are presented in Table 3, and are qualitatively similar to Section 2. The oracle test consistently obtains rejection rates close to the nominal 5% rate across all designs, in accordance with the normality result, even with heterogeneous treatment effects and non-normality of errors due to the binary setup. The L3O rejection rate is close to the nominal rate even with a smaller sample size. EK, MS and MO continue to have high rejection rates in the adversarial designs.

Table 4: Rejection Rates under the null for nominal size 0.05 test for binary X with covariates

	TSLS	EK	MS	MO	JIVEC	ARC	L3O	LMorc
$C_H = S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$	0.048	0.123	0.049	0.052	0.047	0.055	0.054	0.060
$C_H = 2\sqrt{K}, S = 2\sqrt{K}$	0.072	0.111	0.052	0.044	0.041	0.046	0.050	0.053
$C_H = 2\sqrt{K}, S = 2$	0.171	0.016	0.471	0.088	0.012	0.092	0.060	0.050
$C_H = 2\sqrt{K}, S = 0$	0.259	0.002	0.960	0.133	0.008	0.135	0.047	0.058
$C_H = 3, S = 3\sqrt{K}$	0.065	0.132	0.048	0.053	0.056	0.054	0.060	0.049
$C_H = 3, S = 3$	0.131	0.015	0.108	0.040	0.003	0.042	0.044	0.050
$C_H = 3, S = 0$	0.247	0.003	0.300	0.086	0.004	0.091	0.062	0.053
$C_H = 0, S = 2\sqrt{K}$	0.084	0.099	0.054	0.042	0.036	0.043	0.048	0.050
$C_H = 0, S = 2$	0.178	0.006	0.058	0.042	0.002	0.044	0.052	0.051
$C_H = 0, S = 0$	0.246	0.006	0.048	0.063	0.005	0.069	0.081	0.050
$C_H = S = 0, \sigma_{\varepsilon\varepsilon} = 0$	1.000	0.497	0.042	0.015	0.147	0.049	0.092	0.035

Note: Unless stated otherwise, designs use $K = 48, c = 5, \beta = 0, p = 7/8, \sigma_{\varepsilon\varepsilon} = 0.5, \sigma_{\varepsilon v} = 0.1$, and $g = 0.1$ with 1000 simulations. The procedures in the columns are detailed in the remarks of Table 1.

6.3 Incorporating Covariates

This section presents a data-generating process that involves covariates. Instead of judges, consider a model where there are K states. Let $t = 1, \dots, K$ index the state and let W denote the control vector that is an indicator for states. With a binary exogenous variable (say an indicator for birth being in the fourth quarter) $B \in \{0, 1\}$, the value of the instrument is given by $k = t \times B$. Then, the instrument vector Z is an indicator for all possible values of k . The structural model is:

$$Y_i(x) = x(\beta + \xi_i) + w'\gamma + \varepsilon_i, \text{ and}$$

$$X_i(z) = I \{z'\pi + w'\gamma - v_i \geq 0\}.$$

In the simulation, every state has 10 observations, of which 5 have $B = 1$ and the other 5 have $B = 0$. The process for generating $(v_i, \varepsilon_i, \xi_i)$, $\pi_k, \sigma_{\xi vk}$, and s, h is identical to the binary case. Hence, $\pi_0 = \sigma_{\xi v0}$ for the base group, which constitutes half the observations. For $k \neq 0$, π_k is the coefficient for observations from state $t = k$ and have $B = 1$, and $\sigma_{\xi vk}$ is the corresponding heterogeneity term. Whenever $\pi_t = s$, set $\gamma_t = g$; whenever $\pi_t = -s$, set $\gamma_t = -g$. In this setup, it can be shown that the UJIVE estimand is β , and the proof is in the appendix. Table 4 reports the associated simulation results, which are qualitatively similar to the results described before.

7 Conclusion and Extensions

This paper has documented how both weak instruments and heterogeneity can interact to invalidate existing procedures in the many instruments environment. To address both problems simultaneously, this paper has shown that the limiting distribution of commonly-used statistics, including

the LM statistic, in an environment with many weak instruments and heterogeneity, is normal. Further, a leave-three-out variance estimator is consistent for obtaining the variance of the LM statistic. With these two theorems, this paper contributes a feasible method for valid inference. LM is a reasonable procedure in this context, because it is valid and its inversion for confidence sets is simple. The LM test is also reasonably powerful because it has some theoretical optimality properties, and it is also numerically close to the power envelope for local alternatives in the asymptotic experiment, when using a fixed reduced-form covariance matrix calibrated from the [Angrist and Krueger \(1991\)](#) application. However, these results do not imply that the LM procedure is the best or only method for inference. The limiting results on normality and variance estimation are also useful for designing other test procedures, including nearly optimal tests in [Elliott et al. \(2015\)](#), Bonferroni tests that incorporate the reduced-form information, and a VtF procedure adapted to this context, which are potentially more powerful than LM for some alternatives. Exploring potentially more powerful tests is left for future work.

References

- AGAN, A., J. L. DOLEAC, AND A. HARVEY (2023): “Misdemeanor prosecution,” *The Quarterly Journal of Economics*, 138, 1453–1505.
- AIZER, A. AND J. J. DOYLE (2015): “Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges,” *The Quarterly Journal of Economics*, 130, 759–803.
- ANATOLYEV, S. AND M. SØLVSTEN (2023): “Testing many restrictions under heteroskedasticity,” *Journal of Econometrics*, 236, 105473.
- ANDERSON, T. W. AND H. RUBIN (1949): “Estimation of the parameters of a single equation in a complete system of stochastic equations,” *The Annals of mathematical statistics*, 20, 46–63.
- ANDREWS, D. W., M. J. MOREIRA, AND J. H. STOCK (2006): “Optimal two-sided invariant similar tests for instrumental variables regression,” *Econometrica*, 74, 715–752.
- ANDREWS, I., J. H. STOCK, AND L. SUN (2019): “Weak instruments in instrumental variables regression: Theory and practice,” *Annual Review of Economics*, 11, 727–753.
- ANGRIST, J. D., G. W. IMBENS, AND A. B. KRUEGER (1999): “Jackknife instrumental variables estimation,” *Journal of Applied Econometrics*, 14, 57–67.
- ANGRIST, J. D. AND A. B. KRUEGER (1991): “Does compulsory school attendance affect schooling and earnings?” *The Quarterly Journal of Economics*, 106, 979–1014.
- AUTOR, D., A. KOSTØL, M. MOGSTAD, AND B. SETZLER (2019): “Disability benefits, consumption insurance, and household labor supply,” *American Economic Review*, 109, 2613–2654.

- BEKKER, P. A. (1994): “Alternative approximations to the distributions of instrumental variable estimators,” *Econometrica: Journal of the Econometric Society*, 657–681.
- BLANDHOL, C., J. BONNEY, M. MOGSTAD, AND A. TORGOVITSKY (2022): “When is TSLS actually late?” Tech. rep., National Bureau of Economic Research.
- BOOT, T. AND D. NIBBERING (2024): “Inference on LATEs with covariates,” *arXiv preprint arXiv:2402.12607*.
- CATTANEO, M. D., M. JANSSON, AND W. K. NEWHEY (2018): “Inference in linear regression models with many covariates and heteroscedasticity,” *Journal of the American Statistical Association*, 113, 1350–1361.
- CHAO, J. C. AND N. R. SWANSON (2005): “Consistent estimation with a large number of weak instruments,” *Econometrica*, 73, 1673–1692.
- CHAO, J. C., N. R. SWANSON, J. A. HAUSMAN, W. K. NEWHEY, AND T. WOUTERSEN (2012): “Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments,” *Econometric Theory*, 28, 42–86.
- CRUDU, F., G. MELLACE, AND Z. SÁNDOR (2021): “Inference in instrumental variable models with heteroskedasticity and many instruments,” *Econometric Theory*, 37, 281–310.
- DOBBIE, W., J. GOLDIN, AND C. S. YANG (2018): “The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges,” *American Economic Review*, 108, 201–240.
- DONALD, S. G. AND W. K. NEWHEY (2001): “Choosing the number of instruments,” *Econometrica*, 69, 1161–1191.
- DOYLE, J. J. (2007): “Child protection and child outcomes: Measuring the effects of foster care,” *American Economic Review*, 97, 1583–1610.
- DUFOR, J.-M. (1997): “Some impossibility theorems in econometrics with applications to structural and dynamic models,” *Econometrica: Journal of the Econometric Society*, 1365–1387.
- ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2015): “Nearly optimal tests when a nuisance parameter is present under the null hypothesis,” *Econometrica*, 83, 771–811.
- EVDOKIMOV, K. S. AND M. KOLESÁR (2018): “Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects,” *Working Paper*.
- HARTLEY, H. O., J. RAO, AND G. KIEFER (1969): “Variance estimation with one unit per stratum,” *Journal of the American Statistical Association*, 64, 841–851.
- KOLESÁR, M. (2013): “Estimation in an Instrumental Variables Model With Treatment Effect Heterogeneity,” Working Papers 2013-2, Princeton University. Economics Department.

- LEE, D. S., J. MCCRARY, M. J. MOREIRA, AND J. PORTER (2022): “Valid t-ratio Inference for IV,” *American Economic Review*, 112, 3260–3290.
- LEE, D. S., J. MCCRARY, M. J. MOREIRA, J. R. PORTER, AND L. YAP (2023): “What to do when you can’t use ‘1.96’ Confidence Intervals for IV,” Working Paper 31893, National Bureau of Economic Research.
- LEHMANN, E. L. AND J. P. ROMANO (2005): *Testing statistical hypotheses*, vol. 3, Springer.
- LIM, D., W. WANG, AND Y. ZHANG (2024): “A conditional linear combination test with many weak instruments,” *Journal of Econometrics*, 238, 105602.
- MATSUSHITA, Y. AND T. OTSU (2022): “A jackknife Lagrange multiplier test with many weak instruments,” *Econometric Theory*, 1–24.
- MIKUSHEVA, A. AND L. SUN (2022): “Inference with many weak instruments,” *The Review of Economic Studies*, 89, 2663–2686.
- MOREIRA, M. J. (2009a): “A Maximum Likelihood Method for the Incidental Parameter Problem,” *The Annals of Statistics*, 37, 3660–3696.
- (2009b): “Tests with correct size when instruments can be arbitrarily weak,” *Journal of Econometrics*, 152, 131–140.
- PHILLIPS, G. D. A. AND C. HALE (1977): “The bias of instrumental variable estimators of simultaneous equation systems,” *International Economic Review*, 219–228.
- STAIGER, D. AND J. H. STOCK (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65, 557–586.
- STOCK, J. H. AND M. YOGO (2005): “Testing for Weak Instruments in Linear IV Regression,” in *Identification and Inference in Econometric Models: Essays in Honor of Thomas J. Rothenberg*, ed. by D. W. Andrews and J. H. Stock, Cambridge University Press, chap. 5, 80–108.
- YAP, L. (2023): “Valid Wald Inference with Many Weak Instruments,” *arXiv preprint arXiv:2311.15932*.

A Supplementary Material

A.1 Supplementary Results

Lemma 4. Consider the model of Section 2. Suppose $\sigma_{\xi v} \neq 0$, $\sum_k \pi_k = 0$, and $\sum_k \pi_k^2 > 0$. Then, $E[T_{ee}] \neq 0$ for all real β_0 .

Assumption 3 states high-level conditions for consistency of the variance estimator. To ease notation, let R_{mi} stand for either $R_{\Delta i}$ or R_i . Denote $\tilde{R}_i := \sum_{j \neq i} G_{ij} R_j$ and $\tilde{R}_{\Delta i} := \sum_{j \neq i} G_{ij} R_{\Delta j}$. Let $h_4(i, j, k, l)$ be a product of any number of $G_{i_1 i_2}$, $i_1 \neq i_2$, $\tilde{M}_{j_1 j_2}$, $j_1 \neq j_2$, and R_{mk_1} with $i_1, i_2, j_1, j_2, k_1 \in \{i, j, k, l\}$ such that every index in $\{i, j, k, l\}$ occurs at least once as an index of either $G_{i_1 i_2}$ or $\tilde{M}_{j_1 j_2}$. For instance, $h_4(i, j, k, l)$ could be $G_{ij} \tilde{M}_{ik, -il} \tilde{M}_{lj, -ijk}$. Define $h_3(i, j, k)$ and $h_2(i, j)$ in a similar manner. Let $\sum_{i \neq j}^n = \sum_i \sum_{j \neq i}^n$ so that those without the n superscript are still sums of individual indices, but those with an n superscript involves the sum over multiple indices. Objects like $\sum_{i \neq j \neq k}^n$ and $\sum_{i \neq j \neq k \neq l}^n$ are defined in a similar manner. When I refer to the p-sum, I refer to the sum over p non-overlapping indices. For instance, a 3-sum is $\sum_{i \neq j \neq k}^n$. Let F stand for either G or G' .

Assumption 3. For some $C < \infty$,

- (a) $\sum_j G_{ij}^2 \leq C$, $\sum_j G_{ji}^2 \leq C$, $\sum_{j \neq k}^n \left(\sum_{i \neq j, k} G_{ij} F_{ik} \right)^2 \leq \sum_{j \neq k}^n G_{jk}^2$, $\sum_{j \neq k}^n \left(\sum_{i \neq j, k} G_{ji} G_{ki} \right)^2 \leq \sum_{j \neq k}^n G_{jk}^2$, and $|R_{mi}| \leq C$.
- (b) $\sum_{i \neq j \neq k}^n \left(\sum_{l \neq i, j, k} h_4(i, j, k, l) R_{ml} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$, $\sum_{i \neq j}^n \left(\sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4(i, j, k, l) R_{ml} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$, and $\sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4(i, j, k, l) R_{ml} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$.
- (c) $\sum_{i \neq j}^n \left(\sum_{k \neq i, j} h_3(i, j, k) R_{mk} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$ and $\sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} h_3(i, j, k) R_{mk} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$.
- (d) $\sum_i \left(\sum_{j \neq i} h_2(i, j) R_{mj} \right)^2 \leq C \sum_i \tilde{R}_{mi}^2$.

The first condition requires the row and column sums of the squares of the G elements to be bounded. Assumption 1(e) is insufficient because it does not rule out having $G_{ii} = K$ for some i and 0 elsewhere in the G matrix. These remaining conditions can be interpreted as (approximate) sparsity conditions on M and G as the p-sum of entries of \tilde{M} and G cannot be too large. Note that other elements of the covariance matrix can be analogously shown to be consistent using the same strategy by using the appendix lemmas by using \tilde{R}_{Yi} in place of $\tilde{R}_{\Delta i}$ where required.

The judges example in Section 2 satisfies this assumption when there are no covariates and $G = P$ and R values are bounded. For condition (a), $\sum_j P_{ij}^2 = P_{ii} \leq C$ and, since P is idempotent, $\sum_{j \neq k}^n \left(\sum_{i \neq j, k} P_{ij} P_{ik} \right)^2 = \sum_{j \neq k}^n \left(\sum_i P_{ij} P_{ik} - P_{jj} P_{jk} - P_{kk} P_{jk} \right)^2 = \sum_{j \neq k}^n (P_{jk} - P_{jj} P_{jk} - P_{kk} P_{jk})^2 = \sum_{j \neq k}^n (1 - P_{jj} - P_{kk})^2 P_{jk}^2 \leq \sum_{j \neq k}^n P_{jk}^2$. For any \tilde{M}_{ij} and G_{ij} , these elements are nonzero only when i and j share the same judge p . Further, $R_{mi} = \pi_{mp(i)}$, where π_{mp} can denote π_p or $\pi_{\Delta p}$ in the model. Due to how the h functions are defined, when every judge has at most c cases,

$$\begin{aligned} \sum_i \left(\sum_{j \neq i} h_2(i, j) R_{mj} \right)^2 &= \sum_i \left(\sum_{j \in \mathcal{N}_{p(i)} \setminus \{i\}} h_2(i, j) R_{mp(i)} \right)^2 = \sum_p \sum_{i \in \mathcal{N}_p} \left(\sum_{j \in \mathcal{N}_p \setminus \{i\}} h_2(i, j) \pi_{mp} \right)^2 \\ &= \sum_p \sum_{i \in \mathcal{N}_p} \left(\sum_{j \in \mathcal{N}_p \setminus \{i\}} h_2(i, j) \pi_{mp} \right)^2 \pi_{mp}^2 \leq C \sum_p \sum_{i \in \mathcal{N}_p} (c-1)^2 \pi_{mp}^2 = C \sum_i \tilde{R}_{mi}^2. \end{aligned}$$

The same argument applies for the other components. For instance, in other extreme case,

$$\sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4(i, j, k, l) R_{ml} \right)^2 = \sum_p \pi_{mp}^2 \sum_{i \in \mathcal{N}_p} \left(\sum_{j \in \mathcal{N}_p \setminus \{i\}} \sum_{k \in \mathcal{N}_p \setminus \{i, j\}} \sum_{l \in \mathcal{N}_p \setminus \{i, j, k\}} h_4(i, j, k, l) \right)^2$$

$$\leq C \sum_p \sum_{i \in \mathcal{N}_p} \pi_{mp}^2 (c-1)^2 (c-2)^2 (c-3)^2 \leq C \sum_i \tilde{R}_{mi}^2.$$

If there are no covariates so $G = H_Z = P$, as a corollary of Theorem 1, we have the following:

Corollary 1. $E[T_{ee}] = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{\Delta i} R_{\Delta j}$. Further, if $\frac{1}{\sqrt{K}} \sum_i M_{ii} (R_{Yi}^2, R_{Yi} R_i, R_i^2) \rightarrow (C_{YY}, C_Y, C_S)$, then $E[T_{ee}] \rightarrow C_{YY} - 2C_Y \beta + C_S \beta^2$.

With constant treatment effects, $R_{\Delta i} = 0$ so $E[T_{ee}] = 0$, but this no longer holds with heterogeneous treatment effects. Consequently, using the AR statistic without debiasing results in an asymptotic distribution that has nonzero mean. Hence, inference is invalid if we were to naively apply MS22 or [Crudu et al. \(2021\)](#). The [Matsushita and Otsu \(2022\)](#) variance estimator presented in Equation (4) is biased in general. In particular, it can be shown that the model of Section 3.1 implies:

$$\begin{aligned} E[\hat{\Psi}_{MO}] &= \sum_i M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_i M_{ii}^2 R_i^2 E[\nu_i^2] + \sum_i \sum_{j \neq i} P_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E[\eta_j^2] \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} R_j R_{\Delta j} + E[\eta_i \nu_i] R_j R_{\Delta j} + R_i R_{\Delta i} E[\eta_j \nu_j] + E[\eta_i \nu_i] E[\eta_j \nu_j]). \end{aligned} \quad (17)$$

If the R_{Δ} 's are zero, then $\hat{\Psi}_{MO}$ is unbiased, by comparing the expression $E[\hat{\Psi}_{MO}]$ with Equation (9). Heterogeneity results in many excess terms in the expectation of the variance estimator, generating bias and inconsistency in general. However, $\hat{\Psi}_{MO}$ can be consistent when forcing weak identification and weak heterogeneity. If it is assumed that $\frac{1}{\sqrt{K}} \sum_i M_{ii} R_i^2 \rightarrow C_S < \infty$ and $\frac{1}{\sqrt{K}} \sum_i M_{ii} R_{\Delta i}^2 \rightarrow C < \infty$ with weak identification and weak heterogeneity, then the excess terms in $\frac{1}{K} E[\hat{\Psi}_{MO}]$ can be written as $\frac{1}{\sqrt{K}} \frac{1}{\sqrt{K}} \sum_i M_{ii} R_i^2 = \frac{1}{\sqrt{K}} O(1) = o(1)$ and $\frac{1}{\sqrt{K}} \frac{1}{\sqrt{K}} \sum_i M_{ii} R_{\Delta i}^2 = o(1)$. However, when identification or heterogeneity is strong, $\frac{1}{K} \sum_i M_{ii} R_i^2$ or $\frac{1}{K} \sum_i M_{ii} R_{\Delta i}^2$ is nonnegligible and the variance estimator is inconsistent. The variance estimator adapted from MS22 has similar properties. In contrast, the L3O variance estimator is robust regardless of whether the identification is weak or strong.

A.2 More Power Curves

The notation and design of the power curves are identical to Section 4.4. The UW plots (uniform weighting) plots are power curves that put an equal weight on points in the alternative space, in a design similar to Figure 1.

B Derivations for Section 2

B.1 Properties of Model

This subsection derives expressions for relevant objects in the reduced-form model:

$$X_i = Z_i' \pi + \eta_i, \text{ and } Y_i = Z_i' \pi_Y + \zeta_i.$$

Comparing the first-stage equations, $\eta_i = v_i$. As a corollary, for all i , $E[\eta_i^2] = \sigma_{vv}$. Then,

$$Y_i = X_i \beta_i + \varepsilon_i = Z_i' \pi_Y + Z_i' (\pi \beta_i - \pi_Y) + v_i \beta_i + \varepsilon_i.$$

Hence, $\zeta_i = Z_i' (\pi \beta_i - \pi_Y) + v_i \beta_i + \varepsilon_i$. Define π_Y using $E[\zeta_i] = 0$ and $E[v_i \beta_i] = E[v_i (\beta + \xi_i)] = \sigma_{\xi v k(i)}$, which implies:

$$\pi_{Yk} = \pi_k \beta + \sigma_{\xi v k}. \quad (18)$$

Hence, we can rewrite ζ_i as:

$$\zeta_i = \pi_{k(i)} \xi_i - \sigma_{\xi v k(i)} + v_i \beta + v_i \xi_i + \varepsilon_i.$$

Figure 2: One-sided test with $\rho = 0.9$

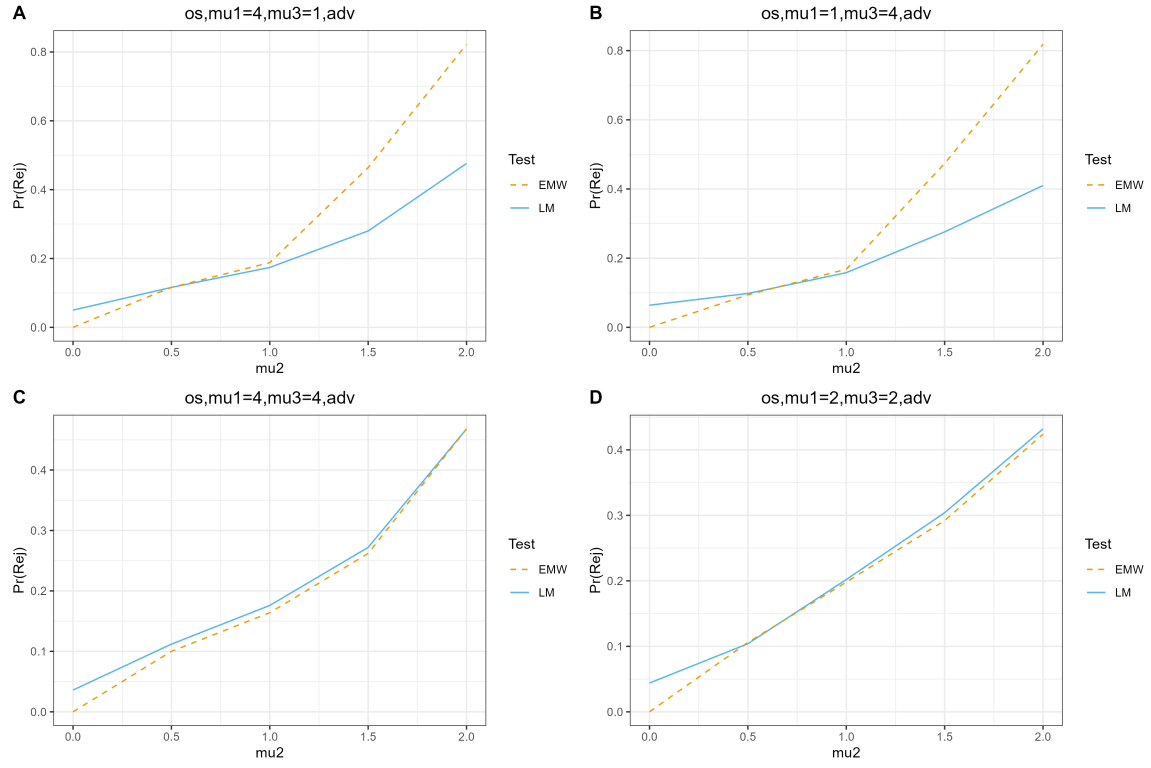


Figure 3: One-sided test with $\rho = 0.34$

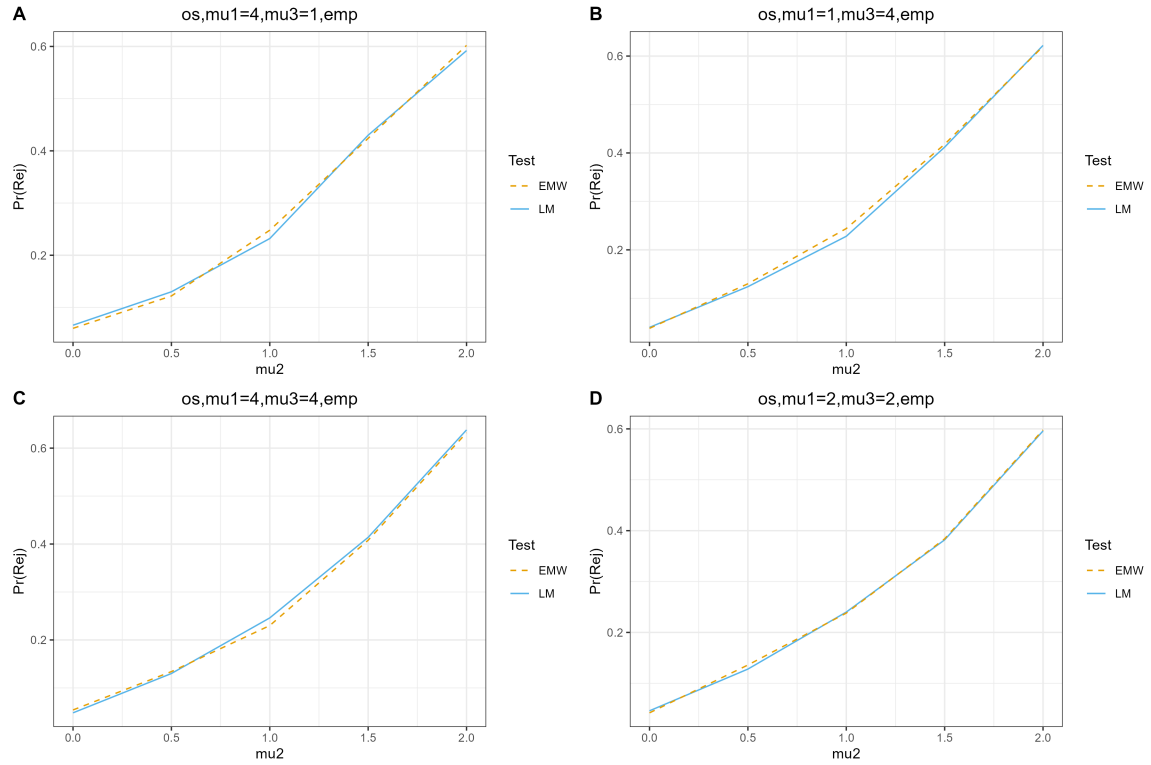


Figure 4: UW with $\rho = 0.9$

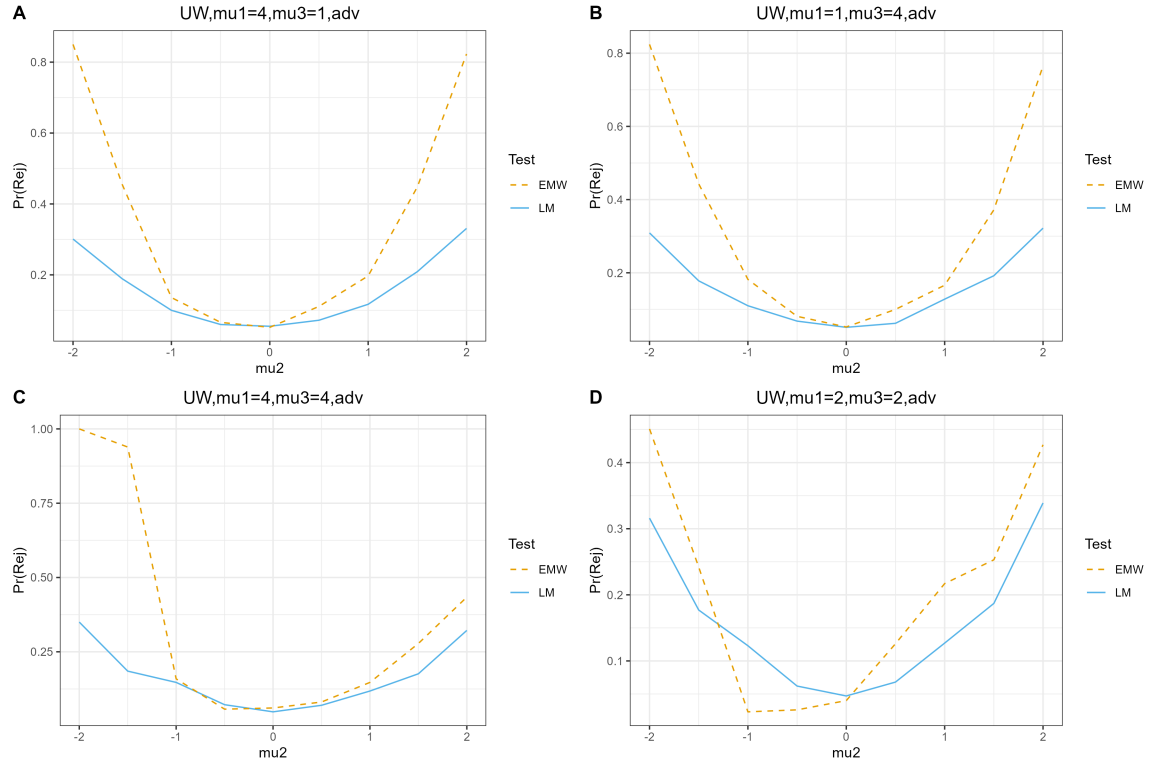
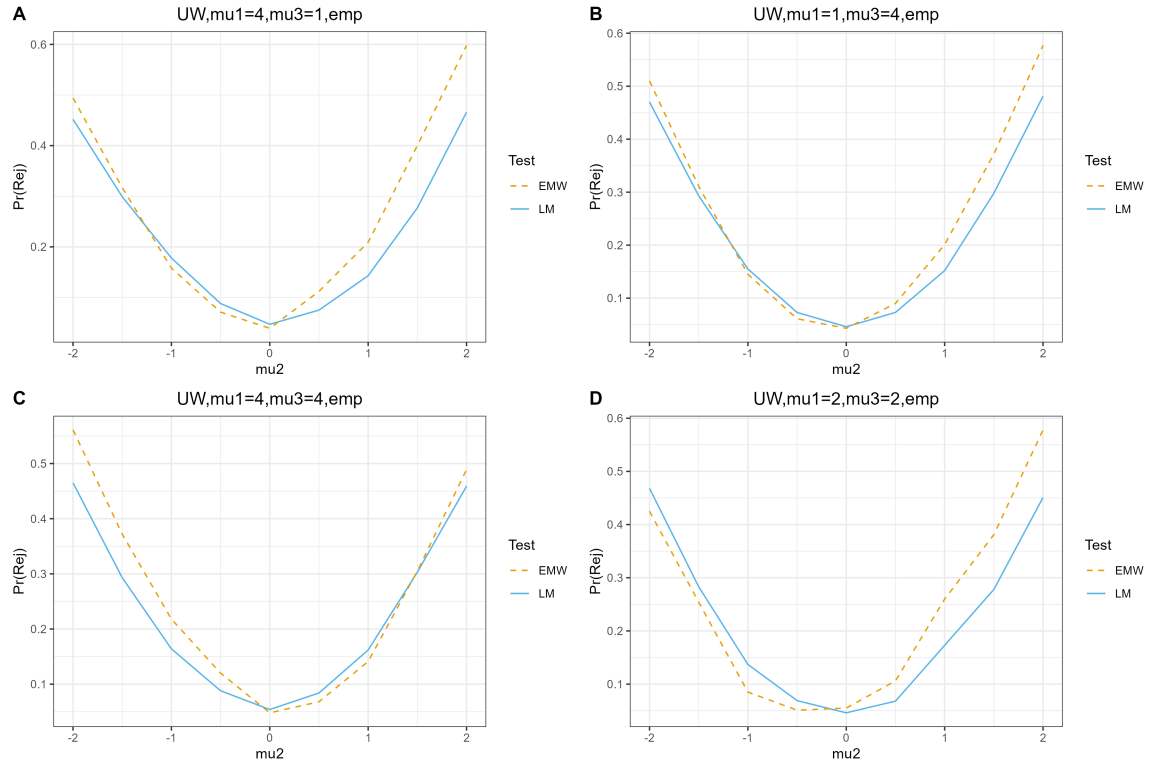


Figure 5: UW with $\rho = 0.34$



By substituting the expression for ζ_i , the covariance is:

$$E[\eta_i \zeta_i | k] = \pi_k \sigma_{\xi vk} + \sigma_{vv} \beta + E[v_i^2 \xi_i] + \sigma_{\varepsilon v}.$$

By Isserlis' theorem, $E[v_i^2 \xi_i] = 0$, so $E[\eta_i \zeta_i | k] = \pi_k \sigma_{\xi vk} + \sigma_{vv} \beta + \sigma_{\varepsilon v}$. Finally, the variance of ζ_i is:

$$\begin{aligned} E[\zeta_i^2 | k] &= E[(\pi_k \xi_i - \sigma_{\xi vk} + v_i \beta_i + \varepsilon_i)^2] \\ &= \pi_k^2 \sigma_{\xi \xi} + \pi_k \beta \sigma_{\xi vk} + \pi_k \sigma_{\varepsilon \xi} + \pi_k \beta \sigma_{\xi vk} - \sigma_{\xi vk}^2 + E[v_i^2 \beta_i^2] + \sigma_{\varepsilon v} \beta + \pi_k \sigma_{\varepsilon \xi} + \sigma_{\varepsilon v} \beta + \sigma_{\varepsilon \varepsilon}. \end{aligned}$$

Since $E[v_i^2 \beta_i^2] = \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi \xi} + 2\sigma_{\xi vk}^2$ by applying Isserlis' theorem, by putting the expressions together, with $\omega_{\eta \eta k} := E[\eta_i^2 | k(i) = k]$, $\omega_{\zeta \eta k} := E[\zeta_i \eta_i | k(i) = k]$, and $\omega_{\zeta \zeta k} := E[\zeta_i^2 | k(i) = k]$, we obtain:

$$\begin{aligned} \omega_{\eta \eta k} &= \sigma_{vv}^2, \\ \omega_{\zeta \eta k} &= \pi_k \sigma_{\xi vk} + \sigma_{vv} \beta + \sigma_{\varepsilon v}, \text{ and} \\ \omega_{\zeta \zeta k} &= \pi_k^2 \sigma_{\xi \xi} + 2\pi_k \beta \sigma_{\xi vk} + 2\pi_k \sigma_{\varepsilon \xi} + \sigma_{\varepsilon \varepsilon} + \sigma_{\xi vk}^2 + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi \xi} + 2\sigma_{\varepsilon v} \beta. \end{aligned} \tag{19}$$

In this model, the local average treatment effect (LATE) of judge k relative to the base judge 0 is:

$$LATE_k = \frac{\pi_{Yk}}{\pi_k} = \beta + \frac{\sigma_{\xi vk}}{\pi_k}. \tag{20}$$

B.2 Derivations with Heterogeneity

Derivation of Equation (3)

$$E[T_{ee}] = \sum_i \sum_{j \neq i} P_{ij} R_{\Delta i} R_{\Delta j} = \sum_k (c-1) \pi_{\Delta k}^2 = \sum_k (c-1) \sigma_{\xi vk}^2$$

The final equality occurs by using $\pi_{Yk} = \pi_k \beta + \sigma_{\xi vk}$ to obtain $\pi_{\Delta k} = \pi_{Yk} - \pi_k \beta = \sigma_{\xi vk}$.

Equation (5) is derived by substituting values in this context into Equation (10).

Derivation of Equation (6) When $\beta = 0$, $\nu_i = \zeta_i$. We can use $M_{ii} = (c-1)/c$, $E[\eta_i^2] = \sigma_{vv}$, and corresponding expressions for $E[\zeta_i \eta_i]$ and $E[\zeta_i^2]$ from Equation (19). Hence, due to Equation (17), and noting that $\sum_i \sum_{j \neq i} P_{ij}^2 = \sum_i \sum_{j \neq i} (1/c^2) = \sum_i \frac{c-1}{c^2} = \sum_k \frac{c-1}{c}$,

$$\begin{aligned} E[\hat{\Psi}_{MO}] &= \sum_i M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_i M_{ii}^2 R_i^2 E[\nu_i^2] + \sum_i \sum_{j \neq i} P_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E[\eta_j^2] \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} R_j R_{\Delta j} + E[\eta_i \nu_i] R_j R_{\Delta j} + R_i R_{\Delta i} E[\eta_j \nu_j] + E[\eta_i \nu_i] E[\eta_j \nu_j]) \\ &= \sum_i \left(\frac{c-1}{c} \right)^2 \pi_{k(i)}^2 \sigma_{\xi vk(i)}^2 + \sum_i \left(\frac{c-1}{c} \right)^2 \pi_{k(i)}^2 \omega_{\zeta \zeta k(i)} + \sum_i \sum_{j \neq i} P_{ij}^2 \omega_{\zeta \zeta k(i)} \sigma_{vv} + \sum_i \sum_{j \neq i} P_{ij}^2 \sigma_{\xi vk(i)} \sigma_{vv} \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 \left(\pi_{k(i)} \pi_{k(j)} \sigma_{\xi vk(i)}^2 + (\pi_{k(i)} \sigma_{\xi vk(i)} + \sigma_{\varepsilon v}) \pi_{k(j)} \sigma_{\xi vk(i)} + \pi_{k(i)} \sigma_{\xi vk(i)} (\pi_{k(j)} \sigma_{\xi vk(i)} + \sigma_{\varepsilon v}) \right) \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 (\pi_{k(i)} \sigma_{\xi vk(i)} + \sigma_{\varepsilon v}) (\pi_{k(j)} \sigma_{\xi vk(i)} + \sigma_{\varepsilon v}) \\ &= \sum_k \frac{(c-1)^2}{c} \pi_k^2 \sigma_{\xi vk}^2 + \sum_k \frac{(c-1)^2}{c} \pi_k^2 \omega_{\zeta \zeta k} + \sum_k \frac{c-1}{c} \omega_{\zeta \zeta k} \sigma_{vv} + \sum_k \frac{c-1}{c} \sigma_{\xi vk} \sigma_{vv} \\ &\quad + \sum_k \frac{c-1}{c} \left(\pi_k^2 \sigma_{\xi vk}^2 + 2(\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v}) \pi_k \sigma_{\xi vk} + (\pi_k \sigma_{\xi vk} + \sigma_{\varepsilon v})^2 \right). \end{aligned}$$

B.3 Derivation for Constructed Instrument

Derivation of Equation (7)

$$\hat{\varepsilon}_i = e_i - \tilde{X}_i \frac{\sum_i e_i \tilde{X}_i}{\sum_i \tilde{X}_i^2} = \frac{e_i \sum_i \tilde{X}_i^2 - \tilde{X}_i \sum_i e_i \tilde{X}_i}{\sum_i \tilde{X}_i^2}$$

$$\begin{aligned} \hat{V} &= \frac{\sum_i \tilde{X}_i^2 \hat{\varepsilon}_i^2}{\left(\sum_i \tilde{X}_i^2\right)^2} = \frac{\sum_i \tilde{X}_i^2 \left(e_i \sum_j \tilde{X}_j^2 - \tilde{X}_i \sum_j e_j \tilde{X}_j\right)^2}{\left(\sum_i \tilde{X}_i^2\right)^4} \\ &= \frac{\sum_i \tilde{X}_i^2 e_i^2 \left(\sum_j \tilde{X}_j^2\right)^2 + \sum_i \tilde{X}_i^4 \left(\sum_j e_j \tilde{X}_j\right)^2 - 2 \sum_i \tilde{X}_i^3 e_i \left(\sum_j \tilde{X}_j^2\right) \left(\sum_j e_j \tilde{X}_j\right)}{\left(\sum_i \tilde{X}_i^2\right)^4} \end{aligned}$$

Applying the asymptotic result that $\frac{1}{n} \sum_j e_j \tilde{X}_j \xrightarrow{p} 0$ from Theorem 1,

$$\begin{aligned} t_{AR}^2 &= \frac{\frac{(\sum_i \tilde{X}_i e_i)^2}{(\sum_i \tilde{X}_i^2)^2}}{\frac{\sum_i \tilde{X}_i^2 e_i^2 (\sum_j \tilde{X}_j^2)^2 + \sum_i \tilde{X}_i^4 (\sum_j e_j \tilde{X}_j)^2 - 2 \sum_i \tilde{X}_i^3 e_i (\sum_j \tilde{X}_j^2) (\sum_j e_j \tilde{X}_j)}{(\sum_i \tilde{X}_i^2)^4}} \\ &= \frac{\left(\frac{1}{\sqrt{n}} \sum_i \tilde{X}_i e_i\right)^2 \left(\frac{1}{n} \sum_i \tilde{X}_i^2\right)^2}{\frac{1}{n} \sum_i \tilde{X}_i^2 e_i^2 \left(\frac{1}{n} \sum_j \tilde{X}_j^2\right)^2 + \frac{1}{n} \sum_i \tilde{X}_i^4 \left(\frac{1}{n} \sum_j e_j \tilde{X}_j\right)^2 - 2 \frac{1}{n} \sum_i \tilde{X}_i^3 e_i \left(\frac{1}{n} \sum_j \tilde{X}_j^2\right) \left(\frac{1}{n} \sum_j e_j \tilde{X}_j\right)} \\ &= \frac{\left(\frac{1}{\sqrt{n}} \sum_i \tilde{X}_i e_i\right)^2 \left(\frac{1}{n} \sum_i \tilde{X}_i^2\right)^2}{\frac{1}{n} \sum_i \tilde{X}_i^2 e_i^2 \left(\frac{1}{n} \sum_j \tilde{X}_j^2\right)^2 + o_P(1)} = \frac{\left(\frac{1}{\sqrt{n}} \sum_i \tilde{X}_i e_i\right)^2}{\frac{1}{n} \sum_i \tilde{X}_i^2 e_i^2} + o_P(1); \text{ and} \end{aligned}$$

$$\begin{aligned} E \left[\sum_i \tilde{X}_i^2 e_i^2 \right] &= \sum_i E \left[\left(\sum_{j \neq i} P_{ij} (R_j + \eta_j) \right)^2 (R_{\Delta i} + \nu_i)^2 \right] = \sum_i E \left[\left(M_{ii}^2 R_i^2 + \left(\sum_{j \neq i} P_{ij}^2 \eta_j^2 \right) \right) (R_{\Delta i}^2 + \nu_i^2) \right] \\ &= \sum_i \left(M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E[\eta_j^2] + M_{ii}^2 R_i^2 E[\nu_i^2] + \sum_{j \neq i} P_{ij}^2 E[\nu_i^2] E[\eta_j^2] \right). \end{aligned}$$

Applying it to this context,

$$E \left[\sum_i \tilde{X}_i^2 e_i^2 \right] = \sum_k \frac{(c-1)^2}{c} \pi_k^2 \sigma_{\xi vk}^2 + \sum_k \frac{c-1}{c} P_{ij}^2 \sigma_{\xi vk}^2 \sigma_{vv} + \sum_k \frac{(c-1)^2}{c} \pi_k^2 \omega_{\zeta \zeta k} + \sum_k \frac{c-1}{c} \omega_{\zeta \zeta k} \sigma_{vv}.$$

C Proofs for Section 3

C.1 Proofs for Section 3.1

First, I prove a quadratic CLT. Let

$$T = \sum_i s'_i v_i + \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j,$$

where v_i is a finite-dimensional random vector independent over $i = 1, \dots, n$ with bounded 4th moments, s_i is a nonstochastic vector that weights the v_i 's, and A is a conformable matrix. Let $\mathbb{I}\{\cdot\}$ denote a function that takes value 1 if the argument is true and 0 if false.

Lemma 5. *Suppose:*

1. $\text{Var}(T)^{-1/2}$ is bounded;
2. $\sum_i s_{il}^4 \rightarrow 0$; and
3. $\|G_L G_L'\|_F + \|G_U G_U'\|_F \rightarrow 0$, where G_L is a lower-triangular matrix with elements $G_{L,ij} = G_{ij} \mathbb{I}\{i > j\}$ and G_U is an upper-triangular matrix with elements $G_{U,ij} = G_{ij} \mathbb{I}\{i < j\}$.

Then, $\text{Var}(T)^{-1/2} T \xrightarrow{d} N(0, 1)$.

Proof of Lemma 5. I rewrite the quadratic term to produce a martingale difference array:

$$\begin{aligned} \sum_i \sum_{j \neq i} G_{ij} v_i' A v_j &= \sum_i \sum_{j < i} G_{ij} v_i' A v_j + \sum_i \sum_{j > i} G_{ij} v_i' A v_j \\ &= \sum_i \sum_{j < i} (G_{ij} v_i' A v_j + G_{ji} v_j' A v_i). \end{aligned}$$

Hence, $\sum_i s_i' v_i + \sum_i \sum_{j \neq i} G_{ij} v_i' A v_j = \sum_i y_i$, where

$$\begin{aligned} y_i &= s_i' v_i + \sum_{j < i} (G_{ij} v_i' A v_j + G_{ji} v_j' A v_i) = s_i' v_i + v_i' A \left(\sum_{j < i} G_{ij} v_j \right) + \left(\sum_{j < i} G_{ji} v_j' \right) A v_i \\ &= s_i' v_i + v_i' A (G_L v)'_{\cdot i} + (G_U v)_i \cdot A v_i. \end{aligned}$$

Let \mathcal{F}_i denote the filtration of y_1, \dots, y_{i-1} . To apply the martingale CLT, we require:

1. $\sum_i E[|y_i|^{2+\epsilon}] \rightarrow 0$.
2. Conditional variance converges to 1, i.e., $P(|\sum_i E[B^2 y_i^2 | \mathcal{F}_i] - 1| > \eta) \rightarrow 0$, where $B = \text{Var}(T)^{-1/2}$.

The 4th moments of v_i are bounded. With $\epsilon = 2$, we want $\sum_i E[y_i^4] \rightarrow 0$. Using Loeve's c_r inequality, it suffices that, for any element l of the v_i vector,

$$\sum_i s_{il}^4 E[v_{il}^4] \rightarrow 0, \text{ and } \sum_i E[v_{il}^4 (G_L v)_{il}^4] \rightarrow 0.$$

The first condition is immediate from condition (2). The second condition holds by condition (3) using the proof in EK18. To be precise,

$$\begin{aligned} \sum_i E[v_{il}^4 (G_L v)_{il}^4] &= \sum_i E[v_{il}^4] E[(G_L v)_{il}^4] \preceq \sum_i E[(G_L v)_{il}^4] \\ &= \sum_i \sum_j G_{L,ij}^4 E[v_{il}^4] + 3 \sum_i \sum_j \sum_{k \neq j} G_{L,ij}^2 G_{L,ik}^2 E[v_{il}^2] E[v_{jl}^2] \\ &\preceq \sum_i \sum_j \sum_k G_{L,ij}^2 G_{L,ik}^2 = \sum_i (G_L G_L')_{ii}^2 \\ &\leq \sum_i \sum_j (G_L G_L')_{ij}^2 = \|G_L G_L'\|_F^2. \end{aligned}$$

The argument for G_U is analogous. Now, I turn to showing convergence of the conditional variance. $y_i = s_i' v_i + v_i' A (G_L v)'_{\cdot i} + (G_U v)_i \cdot A v_i$. With abuse of notation, $W_i = s_i' v_i$ and $X_i = v_i' A (G_L v)'_{\cdot i} + v_i' A (G_U v)'_{\cdot i}$. Since $\text{Var}(BT) = B^2 \sum_i E[W_i^2] + B^2 \sum_i E[X_i^2] = 1$,

$$\sum_i E[B^2 y_i^2 | \mathcal{F}_i] - 1 = B^2 \sum_i (E[X_i^2 | \mathcal{F}_i] - E[X_i^2]) + 2B^2 \sum_i E[W_i X_i | \mathcal{F}_i] + B^2 \sum_i (E[W_i^2 | \mathcal{F}_i] - E[W_i^2]).$$

The previous observations in the filtration do not feature, so $E[W_i^2 | \mathcal{F}_i] - E[W_i^2] = 0$. It suffices to show that the RHS converges to 0. For the $\sum_i E[W_i X_i | \mathcal{F}_i]$ term,

$$\begin{aligned} B^2 \sum_i E[W_i X_i | \mathcal{F}_i] &= B^2 \sum_i E \left[W_i \left(v_i' A (G_L v)'_{i.} + v_i' A (G_U v)'_{i.} \right) | \mathcal{F}_i \right] \\ &= B^2 \sum_i E[W_i v_i' A] (G_L v)'_{i.} + B^2 \sum_i E[W_i v_i' A] (G_U v)'_{i.}. \end{aligned}$$

It suffices to show that the respective squares converge to 0. Due to bounded fourth moments, and applying the Cauchy-Schwarz inequality repeatedly, for some n-vector δ_v with $\|\delta_v\|_2 \leq C$,

$$E \left[\left(\sum_i E[W_i v_i' A] (G_L v)'_{i.} \right)^2 \right] \preceq \delta_v' G_L G_L' \delta_v \leq \|\delta_v\|_2^2 \|G_L G_L'\|_2 \preceq \|G_L G_L'\|_F,$$

and the same argument can be applied to the G_U term. For the other term,

$$\sum_i (E[X_i^2 | \mathcal{F}_i] - E[X_i^2]) = \sum_i \left(E \left[\left(v_i' A (G_L v)'_{i.} + v_i' A (G_U v)'_{i.} \right)^2 | \mathcal{F}_i \right] - E \left[\left(v_i' A (G_L v)'_{i.} + v_i' A (G_U v)'_{i.} \right)^2 \right] \right).$$

It suffices to consider the G_L term, as the G_U and cross terms are analogous:

$$\begin{aligned} &\sum_i \left(E \left[\left(v_i' A (G_L v)'_{i.} \right)^2 | \mathcal{F}_i \right] - E \left[\left(v_i' A (G_L v)'_{i.} \right)^2 \right] \right) \\ &= \sum_i \left((G_L v)_{i.} A' E[v_i v_i'] A (G_L v)'_{i.} - E \left[(G_L v)_{i.} A' v_i v_i' A (G_L v)'_{i.} \right] \right). \end{aligned}$$

Since $\sum_i (G_L v)_{i.} A' E[v_i v_i'] A (G_L v)'_{i.}$ is demeaned, it suffices to show that its variance converges to 0. Due to bounded moments,

$$\text{Var} \left(\sum_i (G_L v)_{i.} A' E[v_i v_i'] A (G_L v)'_{i.} \right) \preceq \sum_i \sum_j (G_L G_L')^2 = \|G_L G_L'\|_F^2,$$

which suffices for the result. □

Proof of Theorem 1. Write the JIVE in terms of reduced-form objects:

$$\begin{aligned} \hat{\beta}_{JIVE} &= \frac{\sum_i \sum_{j \neq i} G_{ij} Y_i X_j}{\sum_i \sum_{j \neq i} G_{ij} X_i X_j} = \frac{\sum_i \sum_{j \neq i} G_{ij} (R_{Yi} + \zeta_i) (R_j + \eta_j)}{\sum_i \sum_{j \neq i} G_{ij} (R_i + \eta_i) (R_j + \eta_j)} \\ &= \frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j + \sum_i \sum_{j \neq i} G_{ij} (\zeta_i R_j + R_{Yi} \eta_j + \zeta_i \eta_j)}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j + \sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j)}. \end{aligned}$$

Use $S^* := \sum_i \sum_{j \neq i} G_{ij} R_i R_j$ to denote the object that is not normalized. Then,

$$\begin{aligned} \hat{\beta}_{JIVE} - \beta_{JIVE} &= \frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j + \sum_i \sum_{j \neq i} G_{ij} (\zeta_i R_j + R_{Yi} \eta_j + \zeta_i \eta_j)}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j + \sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j)} - \frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j} \\ &= \frac{\left(\sum_i \sum_{j \neq i} G_{ij} (\zeta_i R_j + R_{Yi} \eta_j + \zeta_i \eta_j) \right) - \beta \left(\sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j) \right)}{S^* + \sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j)}. \end{aligned}$$

Substitute $\zeta_i = \nu_i + \beta\eta_i$ and $R_{Yi} = R_{\Delta i} - R_i\beta$ in the $\hat{\beta}_{JIVE} - \beta_{JIVE}$ expression to obtain:

$$\hat{\beta}_{JIVE} - \beta_{JIVE} = \frac{\left(\sum_i \sum_{j \neq i} G_{ij} (R_{\Delta i} \eta_j + \nu_i R_j + \nu_i \eta_j)\right)}{S^* + \sum_i \sum_{j \neq i} G_{ij} (R_i \eta_j + R_j \eta_i + \eta_i \eta_j)}.$$

Then, divide by \sqrt{K} to obtain the expression as stated. To see the equivalence with the T objects,

$$\frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} e_i X_j = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (\nu_i R_j + \nu_i \eta_j + R_{\Delta i} R_j + R_{\Delta i} \eta_j),$$

with

$$\begin{aligned} \sum_i \sum_{j \neq i} G_{ij} R_{\Delta i} R_j &= \sum_i \sum_{j \neq i} G_{ij} (R_{Yi} - R_i \beta) R_j \\ &= \sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j - \sum_i \sum_{j \neq i} G_{ij} R_i R_j \left(\frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j} \right) = 0, \end{aligned}$$

while T_{XX} is immediate.

Next, I show that the joint distribution of (T_{ee}, T_{eX}, T_{XX}) is asymptotically normal and derive the mean. Using the Cramer-Wold device, it suffices to show that $c_1 T_{ee} + c_2 T_{eX} + c_3 T_{XX}$ is normal for fixed c 's, where

$$\begin{aligned} c_1 T_{ee} + c_2 T_{eX} + c_3 T_{XX} &= c_1 \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (\nu_i R_j + \nu_i \nu_j + R_{\Delta i} R_{\Delta j} + R_{\Delta i} \nu_j) \\ &\quad + c_2 \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (\nu_i R_j + \nu_i \eta_j + R_{\Delta i} \eta_j) + c_3 \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} (\eta_i R_j + \eta_i \eta_j + R_i R_j + R_i \eta_j). \end{aligned}$$

The object $T = c_1 T_{ee} + c_2 T_{eX} + c_3 T_{XX} - c_1 \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_{\Delta i} R_{\Delta j} - c_3 \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_i R_j$ can be written in the CLT form by setting:

$$\begin{aligned} v_i &= (\eta_i, \nu_i)', \\ s_i &= \begin{bmatrix} c_3 \frac{1}{\sqrt{K}} \sum_{j \neq i} (G_{ij} + G_{ji}) R_j + c_2 \frac{1}{\sqrt{K}} \sum_{j \neq i} G_{ji} R_{\Delta j} \\ c_1 \frac{1}{\sqrt{K}} \sum_{j \neq i} (G_{ij} + G_{ji}) R_{\Delta j} + c_2 \frac{1}{\sqrt{K}} \sum_{j \neq i} G_{ij} R_j \end{bmatrix}, \text{ and} \\ A &= \begin{bmatrix} c_3 & 0 \\ c_2 & c_1 \end{bmatrix}, \end{aligned}$$

so that

$$T = \sum_i s_i' v_i + \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v_i' A v_j.$$

Bounded 4th moments hold by Assumption 1(a). To apply the CLT from Lemma 5, I verify the following:

1. $\text{Var}(T)^{-1/2}$ is bounded;
2. $\sum_i s_{il}^4 \rightarrow 0$ for all l ; and
3. $\|G_L G_L'\|_F + \|G_U G_U'\|_F \rightarrow 0$, where G_L is a lower-triangular matrix with elements $G_{L,ij} = \frac{1}{\sqrt{K}} G_{ij} \mathbb{I}\{i > j\}$ and G_U is an upper-triangular matrix with elements $\frac{1}{\sqrt{K}} G_{U,ij} = G_{ij} \mathbb{I}\{i < j\}$.

Condition (2) follows from Assumption 1(d) and applying the cr-inequality. Condition (3) is immediate from Assumption 1(e). For Condition (1), I show that Assumption 1(b) and (c) imply that, for any non-stochastic scalars c_1, c_2, c_3 that are finite and not all 0, $\text{Var}(c_1 T_{ee} + c_2 T_{eX} + c_3 T_{XX})^{-1/2}$ is bounded. Since

$$\text{Cov} \left(\sum_i s'_i v_i, \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j \right) = 0,$$

$$\text{Var} (T) = \text{Var} \left(\sum_i s'_i v_i \right) + \text{Var} \left(\frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j \right) \geq \text{Var} \left(\frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j \right), \text{ where}$$

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j \right) &= \frac{1}{K} \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E [v'_i A v_j v'_k A v_l] \\ &= \frac{1}{K} \sum_i \sum_{j \neq i} G_{ij}^2 E [v'_i A v_j v'_j A v_i] + \frac{1}{K} \sum_i \sum_{j \neq i} G_{ij} G_{ji} E [v'_i A v_j v'_j A v_i]. \end{aligned}$$

It can be shown that:

$$AE [v_j v'_j] A' = \begin{bmatrix} c_3^2 E [\eta_j^2] & c_2 c_3 E [\eta_j^2] + c_1 c_3 E [\nu_j \eta_j] \\ c_2 c_3 E [\eta_j^2] + c_1 c_3 E [\nu_j \eta_j] & c_2^2 E [\eta_j^2] + 2c_2 c_1 E [\nu_j \eta_j] + c_1^2 E [\nu_j^2] \end{bmatrix}; \text{ and}$$

$$AE [v_j v'_j] A = \begin{bmatrix} c_3^2 E [\eta_j^2] + c_2 c_3 E [\nu_j \eta_j] & c_1 c_3 E [\nu_j \eta_j] \\ c_3 c_2 E [\eta_j^2] + c_3 c_1 E [\nu_j \eta_j] + c_2^2 E [\nu_j \eta_j] + c_2 c_1 E [\nu_j^2] & c_1 c_2 E [\nu_j \eta_j] + c_1^2 E [\nu_j^2] \end{bmatrix}.$$

Hence, for some $\underline{c} > 0$, and $\rho_i := \text{corr}(\nu_i, \eta_i)$,

$$\begin{aligned} \text{Var} \left(\frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} v'_i A v_j \right) &= \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (c_3^2 E [\eta_i^2] E [\eta_j^2] + c_2^2 E [\nu_i^2] E [\eta_j^2] + c_1^2 E [\nu_i^2] E [\nu_j^2]) \\ &\quad + \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (2c_1 c_3 E [\nu_i \eta_i] E [\nu_j \eta_j] + 2c_1 c_2 E [\nu_i^2] E [\nu_j \eta_j] + 2c_2 c_3 E [\eta_i \nu_i] E [\eta_j^2]) \\ &= \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (c_3^2 (1 - \rho_i^2) E [\eta_i^2] E [\eta_j^2] + c_2^2 E [\nu_i^2] E [\eta_j^2] + c_1^2 (1 - \rho_j^2) E [\nu_i^2] E [\nu_j^2]) \\ &\quad + \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (c_3^2 \rho_i^2 E [\eta_i^2] E [\eta_j^2] + c_2^2 E [\nu_i^2] E [\eta_j^2] + c_1^2 \rho_j^2 E [\nu_i^2] E [\nu_j^2]) \\ &\quad + \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \left(2c_1 c_3 \rho_i \sqrt{E [\eta_i^2] E [\nu_i^2]} \rho_j \sqrt{E [\eta_j^2] E [\nu_j^2]} + 2c_1 c_2 E [\nu_i^2] \rho_j \sqrt{E [\eta_j^2] E [\nu_j^2]} \right) \\ &\quad + \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \left(2c_2 c_3 \rho_i \sqrt{E [\eta_i^2] E [\nu_i^2]} E [\eta_j^2] \right) \\ &= \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (c_3^2 (1 - \rho_i^2) E [\eta_i^2] E [\eta_j^2] + c_2^2 E [\nu_i^2] E [\eta_j^2] + c_1^2 (1 - \rho_j^2) E [\nu_i^2] E [\nu_j^2]) \\ &\quad + \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \left(\rho_i c_3 \sqrt{E [\eta_i^2] E [\eta_j^2]} + c_2 \sqrt{E [\nu_i^2] E [\eta_j^2]} + \rho_j c_1 \sqrt{E [\nu_i^2] E [\nu_j^2]} \right)^2 \\ &\geq \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) (c_3^2 (1 - \rho_i^2) E [\eta_i^2] E [\eta_j^2] + c_2^2 E [\nu_i^2] E [\eta_j^2] + c_1^2 (1 - \rho_j^2) E [\nu_i^2] E [\nu_j^2]) \\ &\geq \frac{1}{K} \sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \underline{c} > 0. \end{aligned}$$

The first inequality follows from the observation that $\left(\sum_i \sum_{j \neq i} G_{ij} G_{ji} \right)^2 \leq \left(\sum_i \sum_{j \neq i} G_{ij}^2 \right)^2$ by the

Cauchy-Schwarz inequality, so $\sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij}G_{ji}) \geq 0$. The inequalities in the final line first apply Assumption 1(b), then Assumption 1(c). Finally, since ν_i, η_i are mean zero, the expectations are immediate: $E[T_{ee}] = \sum_i \sum_{j \neq i} G_{ij}R_{\Delta j}R_{\Delta i}$ and $E[T_{XX}] = \sum_i \sum_{j \neq i} G_{ij}R_jR_i$. \square

C.2 Proofs for Section 3.2

Proof of Equation (9). Expanding the variance,

$$\begin{aligned} \text{Var} \left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j \right) &= E \left[\left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j \right)^2 \right] = E \left[\sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} e_i X_j G_{kl} e_k X_l \right] \\ &= \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i X_j \nu_k X_l] + \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i X_j R_{\Delta k} X_l] \\ &\quad + \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[R_{\Delta i} X_j \nu_k X_l] + \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[R_{\Delta i} X_j R_{\Delta k} X_l] \end{aligned}$$

The first term is:

$$\begin{aligned} &\sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i X_j \nu_k X_l] \\ &= \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i R_j \nu_k R_l + \nu_i \eta_j \nu_k R_l + \nu_i R_j \nu_k \eta_l + \nu_i \eta_j \nu_k \eta_l] \\ &= \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i R_j \nu_k R_l + \nu_i \eta_j \nu_k \eta_l] \\ &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\nu_i^2] G_{ij} G_{ik} R_j R_k + \sum_i \sum_{j \neq i} \left(\sum_{l \neq i} G_{ij} G_{il} E[\nu_i \eta_j \nu_i \eta_l] + \sum_{l \neq j} G_{ij} G_{jl} E[\nu_i \eta_j \nu_j \eta_l] \right) \\ &\quad + \sum_i \sum_{j \neq i} \left(\sum_{k \neq i, j} G_{ij} G_{ki} E[\nu_i \eta_j \nu_k \eta_i] + \sum_{k \neq i, j} G_{ij} G_{kj} E[\nu_i \eta_j \nu_k \eta_j] \right) \\ &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\nu_i^2] G_{ij} G_{ik} R_j R_k + \sum_i \sum_{j \neq i} (G_{ij}^2 E[\nu_i^2 \eta_j^2] + G_{ij} G_{ji} E[\nu_i \eta_i \eta_j \nu_j]) \\ &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\nu_i^2] G_{ij} G_{ik} R_j R_k + \sum_i \sum_{j \neq i} (G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + G_{ij} G_{ji} E[\nu_i \eta_i] E[\eta_j \nu_j]) \end{aligned}$$

In the next few terms, the expansion steps are analogous, so intermediate steps are omitted for brevity. The second to fourth terms can be expressed as:

$$\begin{aligned} \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[\nu_i X_j R_{\Delta k} X_l] &= \sum_i E[\nu_i \eta_i] \sum_{j \neq i} G_{ij} R_j \sum_{k \neq i} G_{ki} R_{\Delta k}; \\ \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[R_{\Delta i} X_j \nu_k X_l] &= \sum_i \sum_{j \neq i} \sum_{l \neq i} G_{ji} G_{il} E[\eta_i \nu_i] R_{\Delta j} R_l; \text{ and} \\ \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E[R_{\Delta i} X_j R_{\Delta k} X_l] &= \sum_i E[\eta_i^2] \sum_{j \neq i} \sum_{k \neq i} G_{ji} G_{ki} R_{\Delta j} R_{\Delta k}. \end{aligned}$$

The expression stated in the lemma combines these expressions. The corollary follows from setting $G = P$ and observing that P is symmetric, and that since $PR = I$, we have $\sum_{j \neq i} P_{ij} R_j = \sum_{j \neq i} P_{ji} R_j = M_{ii} R_i$. \square

The proof of Theorem 2 is involved, so it will be split into several intermediate lemmas. First I prove three lemmas that yield useful inequalities, then use the results. The proof strategy of these lemmas is to bound the variances above by components that are in the $h(\cdot)$ form so that Assumption 3 inequalities can

be applied. These inequalities are also sufficiently general that other components of the variance matrix in (8) can be written in the given forms, so repeated applications of these lemmas can analogously show consistency of the associated variance estimators.

Let $V_{mi} = R_{mi} + v_{mi}$ where R_{mi} denotes the nonstochastic component while v_{mi} denotes the mean zero stochastic component. Let $r_n := \sum_i \tilde{R}_i^2 + \sum_i \tilde{R}_{\Delta i}^2 + \sum_i \sum_{j \neq i} G_{ij}^2$. Let C_i, C_{ij}, C_{ijk} any denote nonstochastic objects that are non-negative and are bounded above by C . I use $h_4^A(\cdot)$ and $h_4^B(\cdot)$ to denote two different functions that satisfy the above definition for h_4 .

Lemma 6. *Under Assumption 3, the following hold:*

- (a) $\left| \sum_{i \neq j \neq k} C_{ijk} \left(\sum_{l \neq i, j, k} h_4^A(i, j, k, l) R_{ml} \right) \left(\sum_{l \neq i, j, k} h_4^B(i, j, k, l) R_{ml} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2,$
 $\left| \sum_{i \neq j} C_{ij} \left(\sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4^A(i, j, k, l) R_{ml} \right) \left(\sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4^B(i, j, k, l) R_{ml} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2,$
and $\left| \sum_i C_i \left(\sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4^A(i, j, k, l) R_{ml} \right) \left(\sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} h_4^B(i, j, k, l) R_{ml} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2.$
- (b) $\left| \sum_{i \neq j} C_{ij} \left(\sum_{k \neq i, j} h_3^A(i, j, k) R_{mk} \right) \left(\sum_{k \neq i, j} h_3^B(i, j, k) R_{mk} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2$
and $\left| \sum_i C_i \left(\sum_{j \neq i} \sum_{k \neq i, j} h_3^A(i, j, k) R_{mk} \right) \left(\sum_{j \neq i} \sum_{k \neq i, j} h_3^B(i, j, k) R_{mk} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2.$
- (c) $\left| \sum_i C_i \left(\sum_{j \neq i} h_2^A(i, j) R_{mj} \right) \left(\sum_{j \neq i} h_2^B(i, j) R_{mj} \right) \right| \leq C \sum_i \tilde{R}_{mi}^2.$

Proof of Lemma 6. I begin with part (3). By applying the Cauchy-Schwarz inequality,

$$\begin{aligned}
& \left| \sum_i C_i \left(\sum_{j \neq i} h_2^A(i, j) R_{mj} \right) \left(\sum_{j \neq i} h_2^B(i, j) R_{mj} \right) \right| \\
& \leq \left(\sum_i C_i \left(\sum_{j \neq i} h_2^A(i, j) R_{mj} \right)^2 \right)^{1/2} \left(\sum_i C_i \left(\sum_{j \neq i} h_2^B(i, j) R_{mj} \right)^2 \right)^{1/2} \\
& \leq \max_i C_i \left(\sum_i \left(\sum_{j \neq i} h_2^A(i, j) R_{mj} \right)^2 \right)^{1/2} \left(\sum_i \left(\sum_{j \neq i} h_2^B(i, j) R_{mj} \right)^2 \right)^{1/2} \\
& \leq \max_i C_i \left(\sum_i \tilde{R}_{mi}^2 \right)^{1/2} \left(\sum_i \tilde{R}_{mi}^2 \right)^{1/2} \leq C \sum_i \tilde{R}_{mi}^2.
\end{aligned}$$

The proof of all other parts are entirely analogous. □

Lemma 7. *Under Assumption 3, the following hold:*

- (a) $\text{Var} \left(\sum_{i \neq j} G_{ij} F_{ij} V_{1i} V_{2i} V_{3j} V_{4j} \right) \leq C r_n.$
- (b) $\text{Var} \left(\sum_{i \neq j \neq k} G_{ij} F_{ij} \check{M}_{ik, -ij} V_{1i} V_{2k} V_{3j} V_{4j} \right) \leq C r_n.$
- (c) $\text{Var} \left(\sum_{i \neq j \neq l} G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} V_{3j} V_{4l} \right) \leq C r_n.$
- (d) $\text{Var} \left(\sum_{i \neq j \neq k \neq l} G_{ij} F_{ij} V_{1i} \check{M}_{ik, -ij} V_{2k} V_{3j} \check{M}_{jl, -ijk} V_{4l} \right) \leq C r_n.$

Proof of Lemma 7. Proof of Lemma 7(a).

Using the decomposition from AS23,

$$\text{Var} \left(\sum_i \sum_{j \neq i} G_{ij} F_{ij} V_{1i} V_{2i} V_{3j} V_{4j} \right)$$

$$\begin{aligned}
&= \sum_{i \neq j}^n G_{ij}^2 F_{ij}^2 \text{Var}(V_{1i} V_{2i} V_{3j} V_{4j}) + \sum_{i \neq j}^n G_{ij} F_{ij} G_{ji} F_{ji} \text{Cov}(V_{1i} V_{2i} V_{3j} V_{4j}, V_{1j} V_{2j} V_{3i} V_{4i}) \\
&\quad + \sum_{i \neq j \neq k}^n G_{ij} F_{ij} G_{kj} F_{kj} \text{Cov}(V_{1i} V_{2i} V_{3j} V_{4j}, V_{1k} V_{2k} V_{3j} V_{4j}) + \sum_{i \neq j \neq k}^n G_{ij} F_{ij} G_{jk} F_{jk} \text{Cov}(V_{1i} V_{2i} V_{3j} V_{4j}, V_{1j} V_{2j} V_{3k} V_{4k}) \\
&\quad + \sum_{i \neq j \neq k}^n G_{ij} F_{ij} G_{ik} F_{ik} \text{Cov}(V_{1i} V_{2i} V_{3j} V_{4j}, V_{1i} V_{2i} V_{3k} V_{4k}) + \sum_{i \neq j \neq k}^n G_{ij} F_{ij} G_{ki} F_{ki} \text{Cov}(V_{1i} V_{2i} V_{3j} V_{4j}, V_{1k} V_{2k} V_{3i} V_{4i}) \\
&\leq 2 \left[\max_{i,j} \text{Var}(V_{1i} V_{2i} V_{3j} V_{4j}) \right] \sum_i \left(\left(\sum_{j \neq i} G_{ij} F_{ij} \right)^2 + \left(\sum_{j \neq i} G_{ij} F_{ij} \right) \left(\sum_{j \neq i} G_{ji} F_{ji} \right) \right).
\end{aligned}$$

Notice that the terms in $\sum_{i \neq j}^n$ are absorbed into the sum over k so that the final expression can be written as $\sum_i \sum_{j \neq i} \sum_{k \neq i}$. Then, due to Assumption 3(a) and the Cauchy-Schwarz inequality,

$$\sum_i \left(\sum_{j \neq i} G_{ij} F_{ij} \right)^2 \leq \sum_i \left(\sum_{j \neq i} G_{ij}^2 \right) \left(\sum_{j \neq i} F_{ij}^2 \right) \leq C \sum_i \sum_{j \neq i} G_{ij}^2,$$

and

$$\begin{aligned}
\left| \sum_i \left(\sum_{j \neq i} G_{ij} F_{ij} \right) \left(\sum_{j \neq i} G_{ji} F_{ji} \right) \right| &\leq \left(\sum_i \left(\sum_{j \neq i} G_{ij} F_{ij} \right)^2 \right)^{1/2} \left(\sum_i \left(\sum_{j \neq i} G_{ji} F_{ji} \right)^2 \right)^{1/2} \\
&\leq C \left(\sum_i \sum_{j \neq i} G_{ij}^2 \right)^{1/2} \left(\sum_i \sum_{j \neq i} G_{ji}^2 \right)^{1/2} = C \sum_i \sum_{j \neq i} G_{ij}^2.
\end{aligned}$$

Proof of Lemma 7(b). Expand the term:

$$\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} V_{1i} V_{2k} V_{3j} V_{4j} = \sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} (R_{1i} R_{2k} + v_{1i} R_{2k} + R_{1i} v_{2k} + v_{1i} v_{2k}) V_{3j} V_{4j}.$$

Consider the final sum with 4 stochastic terms. The 6-sums have zero covariances due to independent sampling. The 5-sums also have zero covariances, because at least one of v_1 or v_2 needs to have different indices. Within the 4-sum, the covariance is nonzero only for $j_2 \neq j$. We require i_2 to be equal to either i or k and k_2 the other index. Hence, by bounding covariances above by Cauchy-Schwarz,

$$\begin{aligned}
&\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} v_{1i} v_{2k} V_{3j} V_{4j} \right) \\
&\leq \max_{i,j,k} \text{Var}(v_{1i} v_{2k} V_{3j} V_{4j}) \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} (G_{ij} F_{ij} G_{il} F_{il} \check{M}_{ik, -ij} \check{M}_{ik, -il} + G_{ij} F_{ij} G_{kl} F_{kl} \check{M}_{ik, -ij} \check{M}_{ki, -kl}) \\
&\quad + \max_{i,j,k} \text{Var}(v_{1i} v_{2k} V_{3j} V_{4j}) 3! \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik, -ij}^2 \\
&\leq \max_{i,j,k} \text{Var}(v_{1i} v_{2k} V_{3j} V_{4j}) \left(\sum_{i \neq j \neq k \neq l}^n G_{ij}^2 G_{il}^2 \check{M}_{ik, -ij}^2 \right)^{1/2} \left(\sum_{i \neq j \neq k \neq l}^n F_{ij}^2 F_{il}^2 \check{M}_{ik, -ij}^2 \right)^{1/2}
\end{aligned}$$

$$\begin{aligned}
& + \max_{i,j,k} \text{Var} (v_{1i}v_{2k}V_{3j}V_{4j}) \left(\sum_{i \neq j \neq k \neq l}^n G_{ij}^2 G_{kl}^2 \check{M}_{ik,-ij}^2 \right)^{1/2} \left(\sum_{i \neq j \neq k \neq l}^n F_{ij}^2 F_{kl}^2 \check{M}_{ik,-ij}^2 \right)^{1/2} \\
& + \max_{i,j,k} \text{Var} (v_{1i}v_{2k}V_{3j}V_{4j}) 3! \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik,-ij}^2.
\end{aligned}$$

To obtain the first inequality, observe that once we have fixed 3 indices, there are $3!$ permutations of the $v_{1i}v_{2k}V_{3j}V_{4j}$ that we can calculate covariances for. They are all bounded above by the variance. In the various combinations, we may have different combinations of G and F , but they are bounded above by the expression. To be precise, the 3-sum is:

$$\begin{aligned}
& \sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik,-ij} (G_{ij} F_{ij} \check{M}_{ik,-ij} + G_{ik} F_{ik} \check{M}_{ji,-ik} + G_{ji} F_{ji} \check{M}_{jk,-ji}) \\
& + \sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik,-ij} (G_{jk} F_{jk} \check{M}_{ji,-jk} + G_{ki} F_{ki} \check{M}_{kj,-ki} + G_{kj} F_{kj} \check{M}_{ki,-kj}).
\end{aligned}$$

Apply Cauchy-Schwarz to the sum and apply the commutative property of summations to obtain the upper bound. For instance,

$$\left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik,-ij} G_{jk} F_{jk} \check{M}_{ji,-jk} \right)^2 \leq \left(\sum_{i \neq j \neq k}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik,-ij}^2 \right) \left(\sum_{i \neq j \neq k}^n G_{jk}^2 F_{jk}^2 \check{M}_{ji,-jk}^2 \right).$$

Then, observe that $\sum_i \sum_{j \neq i} \sum_{k \neq i,j} G_{jk}^2 F_{jk}^2 \check{M}_{ji,-jk}^2 = \sum_j \sum_{k \neq j} \sum_{i \neq j,k} G_{jk}^2 F_{jk}^2 \check{M}_{ji,-jk}^2$
 $= \sum_i \sum_{j \neq i} \sum_{k \neq i,j} G_{ij}^2 F_{ij}^2 \check{M}_{ik,-ij}^2$. Due to AS23 Equation (22), $\sum_l \check{M}_{il,-ijk}^2 = O(1)$, so $\sum_{i \neq j \neq k}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik,-ij}^2 \leq C \sum_i \sum_{j \neq i} G_{ij}^2 F_{ij}^2 \leq C \sum_i \sum_{j \neq i} G_{ij}^2$. Similarly, $\sum_{i \neq j \neq k \neq l}^n G_{ij}^2 G_{kl}^2 \check{M}_{ik,-ij}^2 = O(1) \sum_{i \neq j \neq k}^n G_{ij}^2 \check{M}_{ik,-ij}^2 = O(1) \sum_{i \neq j}^n G_{ij}^2$, which delivers the order required.

To deal with 3 stochastic terms,

$$\begin{aligned}
\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik,-ij} R_{1i} v_{2k} V_{3j} V_{4j} \right) &= \text{Var} \left(\sum_{i \neq j}^n v_{2i} V_{3j} V_{4j} \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right) \right) \\
&\leq \sum_{i \neq j} \text{Var} (v_{2i} V_{3j} V_{4j}) \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right) \left[\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} + \sum_{k \neq i,j} G_{ki} F_{ki} \check{M}_{kj,-ki} R_{1k} \right] \\
&+ \max_{i,j} \text{Var} (v_{2i} V_{3j} V_{4j}) \sum_{i \neq j} \sum_{l \neq i,j} \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right) \left(\sum_{k \neq i,l} G_{kl} F_{kl} \check{M}_{ki,-kl} R_{1k} \right) \\
&\leq \sum_i \sum_{j \neq i} \text{Var} (v_{2i} V_{3j} V_{4j}) \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right) \left[\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} + \sum_{k \neq i,j} G_{ki} F_{ki} \check{M}_{kj,-ki} R_{1k} \right] \\
&+ \max_{i,j} \text{Var} (v_{2i} V_{3j} V_{4j}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right)^2 - \max_{i,j} \text{Var} (v_{2i} V_{3j} V_{4j}) \sum_i \sum_{j \neq i} \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right)^2 \\
&\leq \max_{i,j} \text{Var} (v_{2i} V_{3j} V_{4j}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right)^2 \\
&+ \sum_{i \neq j} \text{Var} (v_{2i} V_{3j} V_{4j}) \left(\sum_{k \neq i,j} G_{kj} F_{kj} \check{M}_{ki,-kj} R_{1k} \right) \left(\sum_{k \neq i,j} G_{ki} F_{ki} \check{M}_{kj,-ki} R_{1k} \right)
\end{aligned}$$

To get the first inequality, observe that, if for $l \neq i, j$, we have v_{2l} instead of $V_{3l}V_{4l}$, the covariance must be 0. We can then bound the order by using Assumption 3 and Lemma 6. Similarly,

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} v_{1i} R_{2k} V_{3j} V_{4j} \right) &= \text{Var} \left(\sum_{i \neq j}^n v_{1i} V_{3j} V_{4j} \left(\sum_{k \neq i, j} G_{ij} F_{ij} \check{M}_{ik, -ij} R_{2k} \right) \right) \\ &\leq \max_{i, j} \text{Var} (v_{1i} V_{3j} V_{4j}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ij} F_{ij} \check{M}_{ik, -ij} R_{2k} \right)^2 \\ &\quad + \sum_i \sum_{j \neq i} \text{Var} (v_{1i} V_{3j} V_{4j}) \left(\sum_{k \neq i, j} G_{ij} F_{ij} \check{M}_{ik, -ij} R_{2k} \right) \left(\sum_{k \neq i, j} G_{ji} F_{ji} \check{M}_{jk, -ij} R_{2k} \right). \end{aligned}$$

since the expansion in the intermediate steps are entirely analogous.

Turning to the sum with two stochastic objects,

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} R_{1i} R_{2k} V_{3j} V_{4j} \right) &= \text{Var} \left(\sum_i V_{3i} V_{4i} \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ji} F_{ji} \check{M}_{jk, -ij} R_{1j} R_{2k} \right) \right) \\ &= \sum_i \text{Var} (V_{3i} V_{4i}) \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ji} F_{ji} \check{M}_{jk, -ij} R_{1j} R_{2k} \right)^2 \leq \max_i \text{Var} (V_{3i} V_{4i}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ji} F_{ji} \check{M}_{jk, -ij} R_{1j} R_{2k} \right)^2. \end{aligned}$$

With these inequalities, applying Assumption 3 suffices for the result.

Proof of Lemma 7(c). Expand the term:

$$\sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} V_{3j} V_{4l} = \sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} (R_{3j} R_{4l} + R_{3j} v_{4l} + v_{3j} R_{4l} + v_{3j} v_{4l}).$$

With four stochastic objects,

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} v_{3j} v_{4l} \right) \\ \leq \max_{i, j, k} \text{Var} (V_{1i} V_{2i} v_{3j} v_{4l}) \sum_{i \neq j \neq l}^n \sum_{i_2 \neq i, j, l} (G_{ij} F_{ij} \check{M}_{jl, -ij} G_{i_2 j} F_{i_2 j} \check{M}_{jl, -i_2 j} + G_{ij} F_{ij} \check{M}_{jl, -ij} G_{i_2 l} F_{i_2 l} \check{M}_{lj, -i_2 l}) \\ + \max_{i, j, k} \text{Var} (V_{1i} V_{2i} v_{3j} v_{4l}) 3! \sum_{i \neq j \neq l}^n G_{ij}^2 F_{ij}^2 \check{M}_{jl, -ij}^2. \end{aligned}$$

Simplifying the first line,

$$\begin{aligned} &\sum_{i \neq j \neq l}^n \sum_{i_2 \neq i, j, l} (G_{ij} F_{ij} \check{M}_{jl, -ij} G_{i_2 j} F_{i_2 j} \check{M}_{jl, -i_2 j} + G_{ij} F_{ij} \check{M}_{jl, -ij} G_{i_2 l} F_{i_2 l} \check{M}_{lj, -i_2 l}) \\ &\leq \left(\sum_{i \neq j \neq l \neq i_2}^n G_{ij}^2 G_{i_2 j}^2 \check{M}_{jl, -ij}^2 \right)^{1/2} \left(\sum_{i \neq j \neq l \neq i_2}^n F_{ij}^2 F_{i_2 j}^2 \check{M}_{jl, -i_2 j}^2 \right)^{1/2} \\ &\quad + \left(\sum_{i \neq j \neq l \neq i_2}^n G_{ij}^2 G_{i_2 l}^2 \check{M}_{jl, -ij}^2 \right)^{1/2} \left(\sum_{i \neq j \neq l \neq i_2}^n F_{ij}^2 F_{i_2 l}^2 \check{M}_{lj, -i_2 j}^2 \right)^{1/2}. \end{aligned}$$

These terms have the required order due to a proof analogous to Lemma 7(b). Next,

$$\begin{aligned}
\text{Var} \left(\sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} R_{3j} v_{4l} \right) &= \text{Var} \left(\sum_i \sum_{j \neq i} V_{1i} V_{2i} v_{4j} \left(\sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} \right) \right) \\
&\leq \sum_i \sum_{j \neq i} \text{Var} (V_{1i} V_{2i} v_{4j}) \left(\sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} \right) \left[\sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} + \sum_{l \neq i, j} G_{jl} F_{jl} \check{M}_{li, -jl} R_{3l} \right] \\
&\quad + \max_{i, j} \text{Var} (V_{1i} V_{2i} v_{4j}) \sum_i \sum_{j \neq i} \sum_{i_2 \neq i, j} \left(\sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} \right) \left(\sum_{k \neq i_2, l} G_{kl} F_{kl} \check{M}_{ki_2, -kl} R_{1k} \right) \\
&\leq \max_{i, j} \text{Var} (V_{1i} V_{2i} v_{4j}) \sum_i \left(\sum_{j \neq i} \sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} \right)^2 \\
&\quad + \sum_i \sum_{j \neq i} \text{Var} (V_{1i} V_{2i} v_{4j}) \left(\sum_{l \neq i, j} G_{il} F_{il} \check{M}_{lj, -il} R_{3l} \right) \left(\sum_{l \neq i, j} G_{jl} F_{jl} \check{M}_{li, -jl} R_{3l} \right).
\end{aligned}$$

Further, $\text{Var} \left(\sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} v_{3j} R_{4l} \right)$ can be bounded by a similar argument. Turning to the sum with two stochastic objects,

$$\text{Var} \left(\sum_{i \neq j \neq l}^n G_{ij} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} R_{3j} R_{4l} \right) = \sum_i \text{Var} (V_{1i} V_{2i}) \left(\sum_{j \neq i} \sum_{l \neq i, j} G_{ij} F_{ij} \check{M}_{jl, -ij} R_{3j} R_{4l} \right)^2.$$

These inequalities suffice for the result due to Assumption 3.

Proof of Lemma 7(d). Expand the term:

$$\begin{aligned}
&\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} V_{2k} V_{3j} V_{4l} \\
&= \sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} R_{2k} (R_{3j} R_{4l} + R_{3j} v_{4l} + v_{3j} R_{4l} + v_{3j} v_{4l}) \\
&\quad + \sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} v_{2k} (R_{3j} R_{4l} + R_{3j} v_{4l} + v_{3j} R_{4l} + v_{3j} v_{4l}).
\end{aligned}$$

Consider the v_{2k} line first. We only have the 4-sum to contend with. For 5-sum and above, at least one of the errors can be factored out as a zero expectation. Hence, by using Cauchy-Schwarz and the same argument as above,

$$\begin{aligned}
&\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} v_{2k} v_{3j} v_{4l} \right) \\
&\leq \max_{i, j, k, l} \text{Var} (V_{1i} v_{2k} v_{3j} v_{4l}) 4! \sum_{i \neq j \neq k \neq l}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik, -ij}^2 \check{M}_{jl, -ijk}^2 \\
&\leq C \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ij}^2 \check{M}_{ik, -ij}^2 \leq C \sum_{i \neq j}^n G_{ij}^2 F_{ij}^2 \leq C \left(\sum_{i \neq j}^n G_{ij}^2 \right)^{1/2} \left(\sum_{i \neq j}^n F_{ij}^2 \right)^{1/2}.
\end{aligned}$$

By using the same expansion step as before,

$$\begin{aligned}
& \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ij} \check{M}_{ik, -ij} V_{1i} v_{2k} v_{3j} \left(\sum_{l \neq i, j, k} \check{M}_{jl, -ijk} R_{4l} \right) \right) \\
& \leq \max_{i, j, k} \text{Var} \left(V_{1i} v_{2k} v_{3j} \left(\sum_{l \neq i, j, k} \check{M}_{jl, -ijk} R_{4l} \right) \right) \sum_{i \neq j \neq k \neq i_2}^n (G_{ij} F_{ij} G_{i_2 j} F_{i_2 j} \check{M}_{ik, -ij} \check{M}_{i_2 k, -ij} + G_{ij} F_{ij} G_{i_2 k} F_{i_2 k} \check{M}_{ij, -ik} \check{M}_{i_2 j, -ik}) \\
& \quad + \max_{i, j, k} \text{Var} \left(V_{1i} v_{2k} v_{3j} \left(\sum_{l \neq i, j, k} \check{M}_{jl, -ijk} R_{4l} \right) \right) 3! \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij}^2 F_{ij}^2 \check{M}_{ik, -ij}^2
\end{aligned}$$

The $\sum_{i \neq j \neq k \neq i_2} (G_{ij} F_{ij} G_{i_2 j} F_{i_2 j} \check{M}_{ik, -ij} \check{M}_{i_2 k, -ij} + G_{ij} F_{ij} G_{i_2 k} F_{i_2 k} \check{M}_{ij, -ik} \check{M}_{i_2 j, -ik})$ term has the required order due to the same argument as the proof of Lemma 7(b). Next,

$$\begin{aligned}
& \text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} v_{2k} R_{3j} v_{4l} \right) = \text{Var} \left(\sum_{i \neq j \neq k}^n V_{1i} v_{2k} v_{4j} \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right) \right) \\
& \leq \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_{i \neq j \neq k}^n \sum_{i_2 \neq i, j, k} \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right) \left(\sum_{l \neq i_2, j, k} G_{i_2 l} F_{i_2 l} \check{M}_{i_2 k, -i_2 l} \check{M}_{lj, -i_2 l k} R_{3l} \right) \\
& \quad + \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_{i \neq j \neq k}^n \sum_{i_2 \neq i, j, k} \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right) \left(\sum_{l \neq i_2, j, k} G_{i_2 l} F_{i_2 l} \check{M}_{i_2 j, -i_2 l} \check{M}_{lk, -i_2 l j} R_{3l} \right) \\
& \quad + \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) 3! \sum_{i \neq j \neq k}^n \check{M}_{ik, -ij}^2 \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right)^2 \\
& \leq \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_k \sum_{j \neq k} \left(\sum_{i \neq k, j} \sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right)^2 \\
& \quad - \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_k \sum_{j \neq k} \sum_{i \neq k, j} \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right)^2 \\
& \quad + \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_k \sum_{j \neq k} \left(\sum_{i \neq k, j} \sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right) \left(\sum_{i \neq k, j} \sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ij, -il} \check{M}_{lk, -il j} R_{3l} \right) \\
& \quad - \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) \sum_k \sum_{j \neq k} \sum_{i \neq k, j} \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right) \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ij, -il} \check{M}_{lk, -il j} R_{3l} \right) \\
& \quad + \max_{i, j, k} \text{Var} (V_{1i} v_{2k} v_{4j}) 3! \sum_{i \neq j \neq k}^n \check{M}_{ik, -ij}^2 \left(\sum_{l \neq i, j, k} G_{il} F_{il} \check{M}_{ik, -il} \check{M}_{lj, -ilk} R_{3l} \right)^2.
\end{aligned}$$

The first term in the v_{2k} line is then:

$$\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} v_{2k} R_{3j} R_{4l} \right) = \text{Var} \left(\sum_{i \neq j}^n G_{ij} F_{ij} \check{M}_{ij, -ik} V_{1i} v_{2j} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{kl, -ijk} R_{3k} R_{4l} \right)$$

$$\begin{aligned}
&\leq \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_{i \neq j}^n \left(G_{ij}F_{ij} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right)^2 \\
&+ \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_{i \neq j}^n \left(G_{ij}F_{ij} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right) \left(G_{ji}F_{ji} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ji,-jk} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right) \\
&+ \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_{i \neq j \neq i_2}^n \left(G_{ij}F_{ij} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right) \\
&\quad \left(G_{i_2j}F_{i_2j} \sum_{k \neq i_2,j} \sum_{l \neq i_2,j,k} \check{M}_{i_2j,-i_2k} \check{M}_{kl,-i_2jk} R_{3k} R_{4l} \right) \\
&\leq \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_{i \neq j}^n \left(G_{ij}F_{ij} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right)^2 \\
&+ \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_{i \neq j}^n \left(G_{ij}F_{ij} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right) \left(G_{ji}F_{ji} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \check{M}_{ji,-jk} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right) \\
&+ \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_j \left(\sum_{i \neq j} \sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right)^2 \\
&- \max_{i,j} \text{Var}(V_{1i}v_{2j}) \sum_j \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{ij,-ik} \check{M}_{kl,-ijk} R_{3k} R_{4l} \right)^2.
\end{aligned}$$

Now, we turn back to the R_{2k} expression to complete the proof:

$$\sum_{i \neq j \neq k \neq l}^n G_{ij}F_{ij} \check{M}_{ik,-ij} \check{M}_{jl,-ijk} V_{1i} R_{2k} (R_{3j} R_{4l} + R_{3j} v_{4l} + v_{3j} R_{4l} + v_{3j} v_{4l}).$$

Consider the term with three stochastic terms first, and simplify it using the same strategy as before:

$$\begin{aligned}
&\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij}F_{ij} \check{M}_{ik,-ij} \check{M}_{jl,-ijk} V_{1i} R_{2k} v_{3j} v_{4l} \right) = \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij}F_{ij} V_{1i} v_{3j} v_{4k} \sum_{l \neq i,j,k} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right) \\
&\leq \max_{i,j,k} \text{Var}(V_{1i}v_{3j}v_{4k}) \left(\sum_{k \neq j}^n \left(\sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2 - \sum_{k \neq j}^n \sum_{i \neq k,j} \left(\sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2 \right) \\
&+ \max_{i,j,k} \text{Var}(V_{1i}v_{3j}v_{4k}) \sum_k \sum_{j \neq k} \left(\sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right) \left(\sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ik}F_{ik} \check{M}_{il,-ik} \check{M}_{kj,-ikl} R_{2l} \right) \\
&- \max_{i,j,k} \text{Var}(V_{1i}v_{3j}v_{4k}) \sum_k \sum_{j \neq k} \sum_{i \neq k,j} \left(\sum_{l \neq i,j,k} G_{ij}F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right) \left(\sum_{l \neq i,j,k} G_{ik}F_{ik} \check{M}_{il,-ik} \check{M}_{kj,-ikl} R_{2l} \right) \\
&+ \max_{i,j,k} \text{Var}(V_{1i}v_{3j}v_{4k}) 3! \sum_{i \neq j \neq k}^n \left(G_{ij}F_{ij} \sum_{l \neq i,j,k} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2.
\end{aligned}$$

Next,

$$\begin{aligned}
& \text{Var} \left(\sum_i \sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} R_{2k} v_{3j} R_{4l} \right) \\
& \leq \max_{i, j} \text{Var} (V_{1i} v_{3j}) \sum_{i \neq j} \left(G_{ij} F_{ij} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} R_{2k} R_{4l} \right)^2 \\
& + \max_{i, j} \text{Var} (V_{1i} v_{3j}) \sum_{i \neq j} \left(G_{ij} F_{ij} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} R_{2k} R_{4l} \right) \left(G_{ji} F_{ji} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{jk, -ij} \check{M}_{il, -ijk} R_{2k} R_{4l} \right) \\
& + \max_{i, j} \text{Var} (V_{1i} v_{3j}) \sum_j \left(\sum_{i \neq j} G_{ij} F_{ij} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} R_{2k} R_{4l} \right)^2 \\
& - \max_{i, j} \text{Var} (V_{1i} v_{3j}) \sum_{j \neq i} \left(G_{ij} F_{ij} \sum_{k \neq i, j} \sum_{l \neq i, j, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} R_{2k} R_{4l} \right)^2.
\end{aligned}$$

Finally,

$$\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} R_{2k} R_{3j} R_{4l} \right) = \sum_i \text{Var} (V_{1i}) \left(\sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} G_{ij} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} R_{2k} R_{3j} R_{4l} \right)^2.$$

□

Lemma 8. Under Assumption 3, the following hold:

- (a) $\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{1j} V_{2k} V_{3i} V_{4i} \right) \leq Cr_n.$
- (b) $\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{1j} V_{2k} V_{3i} V_{4l} \right) \leq Cr_n.$

Proof of Lemma 8. Proof of Lemma 8(a). Expand the term:

$$\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{1j} V_{2k} V_{3i} V_{4i} = \sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} V_{4i} (R_{1j} R_{2k} + R_{1j} v_{2k} + v_{1j} R_{2k} + v_{1j} v_{2k}).$$

With four stochastic objects,

$$\begin{aligned}
\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} V_{4i} v_{1j} v_{2k} \right) & \leq \max_{i, j, k} \text{Var} (V_{3i} V_{4i} v_{1j} v_{2k}) \sum_{i \neq j \neq k}^n \sum_{i_2 \neq i, j, k} (G_{ij} F_{ik} G_{i_2 j} F_{i_2 k} + G_{ij} F_{ik} G_{i_2 k} F_{i_2 j}) \\
& + \max_{i, j, k} \text{Var} (V_{1i} V_{2i} v_{3j} v_{4l}) 3! \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ik}^2.
\end{aligned}$$

Observe that, due to Assumption 3(a),

$$\begin{aligned}
\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} G_{lj} F_{lk} & = \sum_{j \neq k}^n \left(\sum_{i \neq j, k} G_{ij} F_{ik} \right) \left(\sum_{l \neq j, k} G_{lj} F_{lk} - G_{ij} F_{ik} \right) \\
& = \sum_{j \neq k}^n \left(\sum_{i \neq j, k} G_{ij} F_{ik} \right)^2 - \sum_{j \neq k \neq i}^n G_{ij}^2 F_{ik}^2
\end{aligned}$$

has the required order, which suffices for the bound. Next,

$$\begin{aligned}
& \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} V_{4i} R_{1j} v_{2k} \right) \\
&= \text{Var} \left(\sum_i \sum_{j \neq i} F_{ij} V_{3i} V_{4i} v_{2j} \left(\sum_{k \neq i, j} G_{ik} R_{1k} \right) \right) \\
&\leq \sum_i \sum_{j \neq i} \text{Var} (V_{3i} V_{4i} v_{2j}) \left(\sum_{k \neq i, j} F_{ij} G_{ik} R_{1k} \right) \left[\sum_{k \neq i, j} F_{ij} G_{ik} R_{1k} + \sum_{k \neq i, j} F_{ji} G_{jk} R_{1k} \right] \\
&+ \max_{i, j} \text{Var} (V_{3i} V_{4i} v_{2j}) \sum_i \sum_{j \neq i} \sum_{i_2 \neq i, j} \left(\sum_{k \neq i, j} F_{ij} G_{ik} R_{1k} \right) \left(\sum_{k \neq i_2, l} F_{i_2 j} G_{i_2 k} R_{1k} \right) \\
&\leq \max_{i, j} \text{Var} (V_{3i} V_{4i} v_{2j}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} F_{ij} G_{ik} R_{1k} \right)^2 + \sum_i \sum_{j \neq i} \text{Var} (V_{3i} V_{4i} v_{2j}) \left(\sum_{k \neq i, j} F_{ij} G_{ik} R_{1k} \right) \left(\sum_{k \neq i, j} F_{ji} G_{jk} R_{1k} \right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} V_{4i} v_{1j} R_{2k} \right) = \text{Var} \left(\sum_i \sum_{j \neq i} V_{3i} V_{4i} v_{1j} \left(\sum_{k \neq i, j} G_{ij} F_{ik} R_{2k} \right) \right) \\
&\leq \max_{i, j} \text{Var} (V_{3i} V_{4i} v_{1j}) \sum_i \left(\left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ij} F_{ik} R_{2k} \right)^2 + \sum_{j \neq i} \left(\sum_{k \neq i, j} G_{ij} F_{ik} R_{2k} \right) \left(\sum_{k \neq i, j} G_{ji} F_{jk} R_{2k} \right) \right)
\end{aligned}$$

Turning to the sum with two stochastic objects,

$$\begin{aligned}
& \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} V_{4i} R_{1j} R_{2k} \right) = \text{Var} \left(\sum_i V_{3i} V_{4i} \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ij} F_{ik} R_{1j} R_{2k} \right) \right) \\
&\leq \max_i \text{Var} (V_{3i} V_{4i}) \sum_i \left(\sum_{j \neq i} \sum_{k \neq i, j} G_{ij} F_{ik} R_{1j} R_{2k} \right)^2
\end{aligned}$$

Proof of Lemma 8(b).

Decompose the term:

$$\begin{aligned}
& \sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{1j} V_{2k} V_{3i} V_{4l} \\
&= \sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} R_{1j} (R_{2k} R_{4l} + R_{2k} v_{4l} + v_{2k} R_{4l} + v_{2k} v_{4l}) \\
&+ \sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} v_{1j} (R_{2k} R_{4l} + R_{2k} v_{4l} + v_{2k} R_{4l} + v_{2k} v_{4l}).
\end{aligned}$$

Consider the v_{1j} line first.

$$\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} v_{1j} v_{2k} v_{4l} \right) \leq \max_{i,j,k,l} \text{Var} (V_{3i} v_{1j} v_{2k} v_{4l}) 4! \sum_{i \neq j \neq k \neq l}^n G_{ij}^2 F_{ik}^2 \check{M}_{il, -ijk}^2.$$

Next, by using the same expansion and simplification steps as before,

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} v_{1j} v_{2k} R_{4l} \right) &= \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} v_{1j} v_{2k} \sum_{l \neq i,j,k} \check{M}_{il, -ijk} R_{4l} \right) \\ &\leq \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{2k}) \sum_k \sum_{j \neq k} \left(\left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il, -ijk} R_{4l} \right)^2 - \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il, -ijk} R_{4l} \right)^2 \right) \\ &\quad + \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{2k}) \sum_k \sum_{j \neq k} \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il, -ijk} R_{4l} \right) \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ik} F_{ij} \check{M}_{il, -ijk} R_{4l} \right) \\ &\quad - \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{2k}) \sum_k \sum_{j \neq k} \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il, -ijk} R_{4l} \right) \left(\sum_{l \neq i,j,k} G_{ik} F_{ij} \check{M}_{il, -ijk} R_{4l} \right) \\ &\quad + \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{2k}) 3! \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ik}^2 \left(\sum_{l \neq i,j,k} \check{M}_{il, -ijk} R_{4l} \right)^2 \end{aligned}$$

and

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} v_{1j} R_{2k} v_{4l} \right) &= \text{Var} \left(\sum_{i \neq j \neq k}^n G_{ij} F_{ik} V_{3i} v_{1j} v_{4k} \sum_{l \neq i,j,k} \check{M}_{ik, -ijl} R_{2l} \right) \\ &\leq \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{4k}) \sum_k \sum_{j \neq k} \left(\left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{ik, -ijl} R_{2l} \right)^2 - \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{ik, -ijl} R_{2l} \right)^2 \right) \\ &\quad + \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{4k}) \sum_k \sum_{j \neq k} \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{ik, -ijl} R_{2l} \right) \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{ik} F_{ij} \check{M}_{il, -ijk} R_{2l} \right) \\ &\quad - \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{4k}) \sum_k \sum_{j \neq k} \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{ik, -ijl} R_{2l} \right) \left(\sum_{l \neq i,j,k} G_{ik} F_{ij} \check{M}_{il, -ijk} R_{2l} \right) \\ &\quad + \max_{i,j,k} \text{Var} (V_{3i} v_{1j} v_{4k}) 3! \sum_{i \neq j \neq k}^n G_{ij}^2 F_{ik}^2 \left(\sum_{l \neq i,j,k} \check{M}_{ik, -ijl} R_{2l} \right)^2 \end{aligned}$$

with $\left(\sum_{l \neq i,j,k} \check{M}_{ik, -ijl} R_{2l} \right)^2 \leq C$. Finally,

$$\begin{aligned} \text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ij} F_{ik} \check{M}_{il, -ijk} V_{3i} v_{1j} R_{2k} R_{4l} \right) &= \text{Var} \left(\sum_{i \neq j}^n G_{ij} V_{3i} v_{1j} \sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ik} \check{M}_{il, -ijk} R_{2k} R_{4l} \right) \\ &\leq \max_{i,j} \text{Var} (V_{3i} v_{1j}) \sum_{i \neq j}^n G_{ij} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ik} \check{M}_{il, -ijk} R_{2k} R_{4l} \right)^2 \end{aligned}$$

$$\begin{aligned}
& + \max_{i,j} \text{Var} (V_{3i}v_{1j}) \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il,-ijk} R_{2k} R_{4l} \right) \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ji} F_{jk} \check{M}_{jl,-ijk} R_{2k} R_{4l} \right) \\
& + \max_{i,j} \text{Var} (V_{3i}v_{1j}) \sum_j \left(\left(\sum_{i \neq j} \sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il,-ijk} R_{2k} R_{4l} \right)^2 - \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il,-ijk} R_{2k} R_{4l} \right)^2 \right)
\end{aligned}$$

Now, return to the R_{1j} line: $\sum_{i \neq j \neq k \neq l} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{3i} R_{1j} (R_{2k} R_{4l} + R_{2k} v_{4l} + v_{2k} R_{4l} + v_{2k} v_{4l})$, so

$$\begin{aligned}
\text{Var} \left(\sum_{i \neq j \neq k \neq l} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{3i} R_{1j} v_{2k} v_{4l} \right) & = \text{Var} \left(\sum_{i \neq j \neq k} G_{il} F_{ik} V_{3i} v_{2k} v_{4j} \sum_{l \neq i,j,k} \check{M}_{ij,-ilk} R_{1l} \right) \\
& \leq \max_{i,j,k} \text{Var} (V_{3i} v_{2k} v_{4j}) \left(\sum_j \sum_{k \neq j} \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{il} F_{ik} \check{M}_{ij,-ilk} R_{1l} \right)^2 - \sum_j \sum_{k \neq j} \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{il} F_{ik} \check{M}_{ij,-ilk} R_{1l} \right)^2 \right) \\
& + \max_{i,j,k} \text{Var} (V_{3i} v_{2k} v_{4j}) \sum_j \sum_{k \neq j} \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{il} F_{ik} \check{M}_{ij,-ilk} R_{1l} \right) \left(\sum_{i \neq j,k} \sum_{l \neq i,j,k} G_{il} F_{ij} \check{M}_{ik,-ilj} R_{1l} \right) \\
& - \max_{i,j,k} \text{Var} (V_{3i} v_{2k} v_{4j}) \sum_j \sum_{k \neq j} \sum_{i \neq j,k} \left(\sum_{l \neq i,j,k} G_{il} F_{ik} \check{M}_{ij,-ilk} R_{1l} \right) \left(\sum_{l \neq i,j,k} G_{il} F_{ij} \check{M}_{ik,-ilj} R_{1l} \right) \\
& + \max_{i,j,k} \text{Var} (V_{3i} v_{2k} v_{4j}) 3! \sum_{i \neq j \neq k} \left(F_{ik} \sum_{l \neq i,j,k} G_{il} \check{M}_{ij,-ilk} R_{1l} \right)^2
\end{aligned}$$

and

$$\begin{aligned}
\text{Var} \left(\sum_{i \neq j \neq k \neq l} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{3i} R_{1j} v_{2k} R_{4l} \right) & = \text{Var} \left(\sum_{i \neq j} F_{ij} V_{3i} v_{2j} \sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ik} \check{M}_{il,-ijk} R_{1k} R_{4l} \right) \\
& \leq \max_{i,j} \text{Var} (V_{3i} v_{2j}) \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ij} G_{ik} \check{M}_{il,-ijk} R_{1k} R_{4l} \right)^2 \\
& + \max_{i,j} \text{Var} (V_{3i} v_{2j}) \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ij} G_{ik} \check{M}_{il,-ijk} R_{1k} R_{4l} \right) \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ij} G_{jk} \check{M}_{jl,-ijk} R_{1k} R_{4l} \right) \\
& + \max_{i,j} \text{Var} (V_{3i} v_{2j}) \sum_j \left(\left(\sum_{i \neq j} \sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ij} G_{ik} \check{M}_{il,-ijk} R_{1k} R_{4l} \right)^2 - \sum_{i \neq j} \left(\sum_{k \neq i,j} \sum_{l \neq i,j,k} F_{ij} G_{ik} \check{M}_{il,-ijk} R_{1k} R_{4l} \right)^2 \right).
\end{aligned}$$

The $\sum_{i \neq j \neq k \neq l} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{3i} R_{1j} R_{2k} v_{4l}$ term is symmetric, because it does not matter which R_m we use. Finally,

$$\text{Var} \left(\sum_{i \neq j \neq k \neq l} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{3i} R_{1j} R_{2k} R_{4l} \right) = \sum_i \text{Var} (V_{3i}) \left(\sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il,-ijk} R_{1j} R_{2k} R_{4l} \right)^2$$

□

Lemma 9. Under Assumption 3, the following hold:

$$(a) \text{Var} \left(\sum_{i \neq j} G_{ji}^2 V_{1i} V_{2i} V_{3j} V_{4j} \right) \leq C r_n;$$

- (b) $\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ji}^2 \check{M}_{ik, -ij} V_{1i} V_{2k} V_{3j} V_{4j} \right) \leq Cr_n;$
- (c) $\text{Var} \left(\sum_{i \neq j \neq l}^n G_{ji}^2 \check{M}_{jl, -ij} V_{1i} V_{2i} V_{3j} V_{4l} \right) \leq Cr_n;$
- (d) $\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ji}^2 V_{1i} \check{M}_{ik, -ij} V_{2k} V_{3j} \check{M}_{jl, -ijk} V_{4l} \right) \leq Cr_n;$
- (e) $\text{Var} \left(\sum_{i \neq j \neq k}^n G_{ji} F_{ki} V_{1j} V_{2k} V_{3i} V_{4i} \right) \leq Cr_n;$
- (f) $\text{Var} \left(\sum_{i \neq j \neq k \neq l}^n G_{ji} F_{ki} \check{M}_{il, -ijk} V_{1j} V_{2k} V_{3i} V_{4l} \right) \leq Cr_n.$

Proof of Lemma 9. The proof of Lemma 9 is entirely analogous to Lemmas 7 and 8 just that G_{ji} is used in place of G_{ij} . \square

Proof of Theorem 2. Proof of Unbiasedness

The variance expression can be equivalently be written as:

$$\begin{aligned} V_{LM} = & \sum_i \left(E[\nu_i^2] \left(\sum_{j \neq i} G_{ij} R_j \right)^2 + 2 \left(\sum_{j \neq i} G_{ij} R_j \right) \left(\sum_{j \neq i} G_{ji} R_{\Delta j} \right) E[\nu_i \eta_i] + E[\eta_i^2] \left(\sum_{j \neq i} G_{ji} R_{\Delta j} \right)^2 \right) \\ & + \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E[\eta_i \nu_i] E[\eta_j \nu_j]. \end{aligned} \quad (21)$$

Take expectation of A_{1i} :

$$\begin{aligned} E \left[\sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ik} X_k e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk}) \right] \\ = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} E[X_j] G_{ik} E[X_k] E[e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk})] + \sum_i \sum_{j \neq i} G_{ij}^2 E[X_j^2] E[e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk})]. \end{aligned}$$

Evaluating the first term,

$$\begin{aligned} \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} E[X_j] G_{ik} E[X_k] E[e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk})] \\ = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} R_j G_{ik} R_k (E[e_i^2] - E[e_i Q'_i \hat{\tau}_{\Delta, -ijk}]) = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} R_j G_{ik} R_k (E[e_i^2] - E[e_i Q'_i \tau_{\Delta}]) \\ = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} R_j G_{ik} R_k E[e_i \nu_i] = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} R_j G_{ik} R_k E[\nu_i^2]. \end{aligned}$$

Using an analogous argument for the second term,

$$\sum_i \sum_{j \neq i} G_{ij}^2 E[X_j^2] E[e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk})] = \sum_i \sum_{j \neq i} G_{ij}^2 (R_j^2 + E[\eta_j^2]) E[\nu_i^2].$$

Combining them,

$$\begin{aligned} E \left[\sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ik} X_k e_i (e_i - Q'_i \hat{\tau}_{\Delta, -ijk}) \right] = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ij} R_j G_{ik} R_k E[\nu_i^2] + \sum_i \sum_{j \neq i} G_{ij}^2 (R_j^2 + E[\eta_j^2]) E[\nu_i^2] \\ = \sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ij} R_j G_{ik} R_k E[\nu_i^2] + \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2]. \end{aligned}$$

Similarly,

$$E[A_{2i}] = \left(\sum_{j \neq i} G_{ij} R_j \right) \left(\sum_{j \neq i} G_{ji} R_{\Delta j} \right) E[\nu_i \eta_i] + \sum_{j \neq i} G_{ij} G_{ji} E[\eta_i \nu_i] E[\eta_j \nu_j], \text{ and}$$

$$E[A_{3i}] = E[\eta_i^2] \left(\sum_{j \neq i} G_{ji} R_{\Delta j} \right)^2 + \sum_{j \neq i} G_{ji}^2 E[\eta_i^2] E[\nu_j^2].$$

For the A4 and A5 terms, observe that:

$$\begin{aligned} X_i - Q'_i \hat{\tau}_{-ij} &= X_i - Q'_i \sum_{k \neq i, j} \left(\sum_{l \neq i, j} Q_l Q'_l \right)^{-1} Q_k X_k \\ &= X_i + \sum_{k \neq i, j} \check{M}_{ik, -ij} X_k = \sum_{k \neq j} \check{M}_{ik, -ij} X_k, \end{aligned}$$

where the final equality follows from $\check{M}_{ii, -ij} = 1$. Then,

$$\begin{aligned} E[A_{4ij}] &= E \left[X_i \sum_{k \neq j} \check{M}_{ik, -ij} X_k e_j (X_j - Q'_j \hat{\tau}_{\Delta, -ijk}) \right] = \sum_{k \neq j} E[X_i \check{M}_{ik, -ij} X_k e_j (X_j - Q'_j \hat{\tau}_{\Delta, -ijk})] \\ &= E \left[X_i \sum_{k \neq j} \check{M}_{ik, -ij} X_k \right] E[e_j (e_j - Q'_j \hat{\tau}_{\Delta, -ijk})] = E[X_i (X_i - Q'_i \hat{\tau}_{-ij})] E[\nu_j^2] = E[\eta_i^2] E[\nu_j^2]. \end{aligned}$$

Similarly, $E[A_{5ij}] = E[\eta_i \nu_i] E[\eta_j \nu_j]$. Combining these expressions yields the unbiasedness result.

Proof of Consistency

By Chebyshev's inequality,

$$\begin{aligned} &Pr \left(\left| \frac{\hat{V}_{LM} - \text{Var} \left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j \right)}{\text{Var} \left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j \right)} \right| > \epsilon \right) \\ &\leq \frac{1}{\epsilon^2} \frac{\text{Var} \left(\sum_i (A_{1i} + 2A_{2i} + A_{3i}) - \sum_i \sum_{j \neq i} G_{ji}^2 A_{4ij} - \sum_i \sum_{j \neq i} G_{ij} G_{ji} A_{5ij} \right)}{\left[\text{Var} \left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j \right) \right]^2} \end{aligned}$$

Observe that the numerator can be written as the variance of the estimator only because \hat{V}_{LM} is unbiased. I first establish the order of the denominator. As in the supplement, denote $\tilde{R}_i := \sum_{j \neq i} G_{ij} R_j$ and $\tilde{R}_{\Delta i} := \sum_{j \neq i} G_{ji} R_{\Delta j}$. Further, to simplify notation, let $\rho_i := \text{corr}(\eta_i \nu_i)$.

Since $E[\nu_i^2]$ and $E[\eta_i^2]$ are bounded away from zero and $|\text{corr}(\eta_i \nu_i)|$ is bounded away from one by Assumption 1(b), the first line of the V_{LM} expression in Equation (21) has order at least $\sum_i \tilde{R}_i^2 + \sum_i \tilde{R}_{\Delta i}^2$, and the second line has order at least $\sum_i \sum_{j \neq i} G_{ij}^2$. To see this, for some $\underline{c} > 0$, the first line is:

$$\begin{aligned} &\sum_i E[\nu_i^2] \tilde{R}_i^2 + 2\tilde{R}_{\Delta i} \tilde{R}_i E[\nu_i \eta_i] + \tilde{R}_{\Delta i}^2 E[\eta_i^2] = \sum_i E[\nu_i^2] \tilde{R}_i^2 + 2\tilde{R}_{\Delta i} \tilde{R}_i \rho_i \sqrt{E[\nu_i^2] E[\eta_i^2]} + \tilde{R}_{\Delta i}^2 E[\eta_i^2] \\ &\geq \sum_i \left(E[\nu_i^2] \tilde{R}_i^2 + \tilde{R}_{\Delta i}^2 E[\eta_i^2] \right) (1 - |\rho_i|) + \sum_i |\rho_i| \left(E[\nu_i^2] \tilde{R}_i^2 + \tilde{R}_{\Delta i}^2 E[\eta_i^2] - 2\tilde{R}_{\Delta i} \tilde{R}_i \sqrt{E[\nu_i^2] E[\eta_i^2]} \right) \\ &= \sum_i \left(E[\nu_i^2] \tilde{R}_i^2 + \tilde{R}_{\Delta i}^2 E[\eta_i^2] \right) (1 - |\rho_i|) + \sum_i |\rho_i| \left(\sqrt{E[\nu_i^2] \tilde{R}_i^2} - \sqrt{\tilde{R}_{\Delta i}^2 E[\eta_i^2]} \right)^2 \end{aligned}$$

$$\geq \sum_i \left(E[\nu_i^2] \tilde{R}_i^2 + \tilde{R}_{\Delta i}^2 E[\eta_i^2] \right) (1 - |\rho_i|) \geq \underline{c} \sum_i \left(\tilde{R}_i^2 + \tilde{R}_{\Delta i}^2 \right),$$

and the second line is:

$$\begin{aligned} & \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E[\eta_i \nu_i] E[\eta_j \nu_j] \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E[\eta_i \nu_i] E[\eta_j \nu_j] + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 E[\nu_j^2] E[\eta_i^2] \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} \rho_i \rho_j \sqrt{E[\nu_i^2] E[\eta_j^2]} \sqrt{E[\nu_j^2] E[\eta_i^2]} + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 E[\nu_j^2] E[\eta_i^2] \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] (1 - \rho_i^2) + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 E[\nu_j^2] E[\eta_i^2] (1 - \rho_j^2) \\ &\quad + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] \rho_i^2 + \sum_i \sum_{j \neq i} G_{ij} G_{ji} \rho_i \rho_j \sqrt{E[\nu_i^2] E[\eta_j^2]} \sqrt{E[\nu_j^2] E[\eta_i^2]} + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 \rho_j^2 E[\nu_j^2] E[\eta_i^2] \\ &= \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] (1 - \rho_i^2) + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 E[\nu_j^2] E[\eta_i^2] (1 - \rho_j^2) \\ &\quad + \frac{1}{2} \sum_i \sum_{j \neq i} \left(G_{ij} \rho_i \sqrt{E[\nu_i^2] E[\eta_j^2]} + G_{ji} \rho_j \sqrt{E[\nu_j^2] E[\eta_i^2]} \right)^2 \\ &\geq \frac{1}{2} \sum_i \sum_{j \neq i} G_{ij}^2 E[\nu_i^2] E[\eta_j^2] (1 - \rho_i^2) + \frac{1}{2} \sum_i \sum_{j \neq i} G_{ji}^2 E[\nu_j^2] E[\eta_i^2] (1 - \rho_j^2) \geq \underline{c} \sum_i \sum_{j \neq i} G_{ij}^2. \end{aligned}$$

Consequently,

$$V_{LM} \succeq \sum_i \tilde{R}_i^2 + \sum_i \tilde{R}_{\Delta i}^2 + \sum_i \sum_{j \neq i} G_{ij}^2 =: r_n. \quad (22)$$

Due to Assumption 1(c), $\sum_i \sum_{j \neq i} (G_{ij}^2 + G_{ij} G_{ji}) \succeq K$. Since $\left(\sum_i \sum_{j \neq i} G_{ij} G_{ji} \right)^2 \leq \left(\sum_i \sum_{j \neq i} G_{ij}^2 \right)^2$, $\sum_i \sum_{j \neq i} G_{ij}^2 \succeq K \rightarrow \infty$. Hence, V_{LM} diverges, as $r_n \rightarrow \infty$. By repeated application of the Cauchy-Schwarz inequality, it suffices to show that the variance of each of the 5 A terms above has order at most r_n (i.e., bounded by any of the three terms in Equation (22)). If this is true, then since the denominator has order at least r_n^2 , the variance estimator is consistent. Since the derivations are analogous, I focus on $\text{Var}(\sum_i A_{1i})$ and $\text{Var}(\sum_i \sum_{j \neq i} G_{ji}^2 A_{4ij})$. The A1 and A2 terms have the form:

$$\begin{aligned} & \sum_i \sum_{j \neq i} G_{ij} F_{ik} V_{1j} \sum_{k \neq i} V_{2k} V_{3i} (V_{4i} - Q'_i \hat{\tau}_{4,-ijk}) = \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq j,k} G_{ij} F_{ik} V_{1j} V_{2k} V_{3i} \check{M}_{il,-ijk} V_{4l} \\ &= \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} G_{ij} F_{ik} \check{M}_{il,-ijk} V_{1j} V_{2k} V_{3i} V_{4l} + \sum_i \sum_{j \neq i} \sum_{k \neq i,j} G_{ij} F_{ik} V_{1j} V_{2k} V_{3i} V_{4i} \\ &\quad + \sum_i \sum_{j \neq i} \sum_{l \neq i,j} G_{ij} F_{ij} \check{M}_{il,-ij} V_{1j} V_{2j} V_{3i} V_{4l} + \sum_i \sum_{j \neq i} G_{ij} F_{ij} V_{1j} V_{2j} V_{3i} V_{4i}. \end{aligned}$$

In particular, A1 uses $F = G, V_1 = X, V_2 = X, V_3 = e, V_4 = e$, while A2 uses $F = G', V_1 = X, V_2 = e, V_3 = e, V_4 = X$. By applying the Cauchy-Schwarz inequality, it suffices to show that the variance of each of the sums has order at most r_n . The terms $\sum_i \sum_{j \neq i} G_{ij} F_{ij} V_{1j} V_{2j} V_{3i} V_{4i}$ and $\sum_i \sum_{j \neq i} \sum_{l \neq i,j} G_{ij} F_{ij} \check{M}_{il,-ij} V_{1j} V_{2j} V_{3i} V_{4l}$ are identical to the result in Lemma 7, with the latter result being obtained by switching the i and j indices. The remaining terms have a variance that has a bounded order by Lemma 8. For A3, we can use G_{ji} in place of G_{ij} above, and use $F = G', V_1 = e, V_2 = e, V_3 = X, V_4 = X$ so that the order is bounded above due

to Lemma 9. A4 and A5 can be written as:

$$\begin{aligned}
& \sum_i \sum_{j \neq i} G_{ji} F_{ij} V_{1i} \sum_{k \neq j} \check{M}_{ik, -ij} V_{2k} V_{3j} (V_{4j} - Q'_j \hat{\tau}_{4, -ijk}) = \sum_i \sum_{j \neq i} \sum_{k \neq j} \sum_{l \neq i, k} G_{ji} F_{ij} V_{1i} \check{M}_{ik, -ij} V_{2k} V_{3j} \check{M}_{jl, -ijk} V_{4l} \\
& = \sum_i \sum_{j \neq i} \sum_{k \neq i, j} \sum_{l \neq i, j, k} G_{ji} F_{ij} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} V_{1i} V_{2k} V_{3j} V_{4l} + \sum_i \sum_{j \neq i} \sum_{k \neq i, j} G_{ji} F_{ij} \check{M}_{ik, -ij} V_{1i} V_{2k} V_{3j} V_{4j} \\
& + \sum_i \sum_{j \neq i} \sum_{l \neq i, j} G_{ji} F_{ij} \check{M}_{jl, -ij} V_{1i} V_{2i} V_{3j} V_{4l} + \sum_i \sum_{j \neq i} G_{ji} F_{ij} V_{1i} V_{2i} V_{3j} V_{4j}.
\end{aligned}$$

In particular, A4 uses $F = G', V_1 = X, V_2 = X, V_3 = e, V_4 = e$, while A5 uses $F = G, V_1 = e, V_2 = X, V_3 = e, V_4 = X$. By applying the Cauchy-Schwarz inequality, it suffices to show that the variance of each of the sums has order at most r_n . This result is immediate from Lemma 7 and Lemma 9. \square

D Proofs for Section 4

D.1 Proofs for Section 4.1

Proof of Lemma 1. The joint distribution of $(Y', X')'$ is:

$$\begin{bmatrix} Y \\ X \end{bmatrix} \sim N \left(\begin{bmatrix} Z\pi_Y \\ Z\pi \end{bmatrix}, \begin{bmatrix} I_n \omega_{\zeta\zeta} & I_n \omega_{\zeta\eta} \\ I_n \omega_{\eta\zeta} & I_n \omega_{\eta\eta} \end{bmatrix} \right).$$

Stack them together with their predicted values $PY = Z(Z'Z)^{-1}Z'Y$ and $PX = Z(Z'Z)^{-1}Z'X$:

$$\begin{bmatrix} Y \\ X \\ Z(Z'Z)^{-1}Z'Y \\ Z(Z'Z)^{-1}Z'X \end{bmatrix} \sim N \left(\begin{bmatrix} Z\pi_Y \\ Z\pi \end{bmatrix}, \begin{bmatrix} I_n \omega_{\zeta\zeta} & I_n \omega_{\zeta\eta} & \omega_{\zeta\zeta} Z(Z'Z)^{-1}Z' & \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' \\ I_n \omega_{\eta\zeta} & I_n \omega_{\eta\eta} & \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' & \omega_{\eta\eta} Z(Z'Z)^{-1}Z' \\ \omega_{\zeta\zeta} Z(Z'Z)^{-1}Z' & \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' & \omega_{\zeta\zeta} Z(Z'Z)^{-1}Z' & \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' \\ \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' & \omega_{\eta\eta} Z(Z'Z)^{-1}Z' & \omega_{\zeta\eta} Z(Z'Z)^{-1}Z' & \omega_{\eta\eta} Z(Z'Z)^{-1}Z' \end{bmatrix} \right).$$

Then, the conditional normal distribution is:

$$\begin{aligned}
\begin{bmatrix} Y \\ X \end{bmatrix} \mid \begin{bmatrix} Z(Z'Z)^{-1}Z'Y \\ Z(Z'Z)^{-1}Z'X \end{bmatrix} & \sim N \left(\begin{bmatrix} Z\pi_Y \\ Z\pi \end{bmatrix} + \begin{bmatrix} Z(Z'Z)^{-1}Z'Y - Z\pi_Y \\ Z(Z'Z)^{-1}Z'X - Z\pi \end{bmatrix}, V \right) \\
& = N \left(\begin{bmatrix} Z(Z'Z)^{-1}Z'Y \\ Z(Z'Z)^{-1}Z'X \end{bmatrix}, V \right) = N \left(\begin{bmatrix} PY \\ PX \end{bmatrix}, V \right)
\end{aligned}$$

Hence, PX and PY (i.e, $Z'X, Z'Y$) are sufficient statistics for π_Y, π .

To show that $(s'_1 s_1, s'_1 s_2, s'_2 s_2)$ is a maximal invariant, let F be some conformable orthogonal matrix so $F'F = I$. For invariance, let $s_1^* = Fs_1$. Then, $s_1^{*'} s_1^* = s_1' F' F s_1 = s_1' s_1$. Invariance of $(s'_1 s_2, s'_2 s_2)$ is analogous. Maximality states that if $s_1^{*'} s_1^* = s_1' s_1$, then $s_1^* = Fs_1$ for some F . Suppose not. This means $s_1^* = Gs_1$, and G is not an orthogonal matrix but yet $s_1^{*'} s_1^* = s_1' s_1$. Since G is not an orthogonal matrix, $G'G \neq I$. Hence, $s_1^{*'} s_1^* = s_1' G' G s_1 \neq s_1' s_1$, a contradiction. To obtain the distribution,

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} (Z'Z)^{-1/2} Z' (Z\pi_Y + \zeta) \\ (Z'Z)^{-1/2} Z' (Z\pi + \eta) \end{bmatrix} = \begin{bmatrix} (Z'Z)^{1/2} \pi_Y \\ (Z'Z)^{1/2} \pi \end{bmatrix} + \begin{bmatrix} (Z'Z)^{-1/2} Z' \zeta \\ (Z'Z)^{-1/2} Z' \eta \end{bmatrix}.$$

Since $\text{Var} \left((Z'Z)^{-1/2} Z' \eta \right) = (Z'Z)^{-1/2} Z' \omega_{\eta\eta} Z (Z'Z)^{-1/2} = I_K \omega_{\eta\eta}$,

$$\begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \sim N \left(\begin{bmatrix} (Z'Z)^{1/2} \pi_Y \\ (Z'Z)^{1/2} \pi \end{bmatrix}, \Omega \otimes I_K \right).$$

\square

Proof of Proposition 1. Let

$$\begin{pmatrix} \Pi_Y \\ \Pi \end{pmatrix} := \begin{pmatrix} (Z'Z)^{1/2} \pi_Y \\ (Z'Z)^{1/2} \pi \end{pmatrix}.$$

With this definition, $(\pi_Y' Z' Z \pi_Y, \pi_Y' Z' Z \pi_Y, \pi_Y' Z' Z \pi) = (\Pi_Y' \Pi_Y, \Pi_Y' \Pi, \Pi' \Pi)$, and

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \Pi_Y \\ \Pi \end{pmatrix}, \Omega \otimes I_K \right).$$

Split s_1 and s_2 into the Π component and a random normal component: $s_{1k} = \Pi_{Yk} + z_{1k}$ and $s_{2k} = \Pi_k + z_{2k}$. Then, for all k ,

$$\begin{pmatrix} z_{1k} \\ z_{2k} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \omega_{\zeta\zeta} & \omega_{\zeta\eta} \\ \omega_{\zeta\eta} & \omega_{\eta\eta} \end{bmatrix} \right), \text{ and}$$

$$\begin{aligned} \begin{pmatrix} s_1' s_1 \\ s_1' s_2 \\ s_2' s_2 \end{pmatrix} &= \begin{pmatrix} \sum_k s_{1k}^2 \\ \sum_k s_{1k} s_{2k} \\ \sum_k s_{2k}^2 \end{pmatrix} = \begin{pmatrix} \sum_k (\Pi_{Yk} + z_{1k})^2 \\ \sum_k (\Pi_{Yk} + z_{1k})(\Pi_k + z_{2k}) \\ \sum_k (\Pi_k + z_{2k})^2 \end{pmatrix} \\ &= \begin{pmatrix} \sum_k \Pi_{Yk}^2 + 2 \sum_k \Pi_{Yk} z_{1k} + \sum_k z_{1k}^2 \\ \sum_k \Pi_{Yk} \Pi_k + \sum_k \Pi_{Yk} z_{2k} + \sum_k \Pi_k z_{1k} + \sum_k z_{1k} z_{2k} \\ \sum_k \Pi_k^2 + 2 \sum_k \Pi_k z_{2k} + \sum_k z_{2k}^2 \end{pmatrix}. \end{aligned}$$

Under the assumption, $\Pi' \Pi / \sqrt{K} \rightarrow C_S$, so $\frac{1}{\sqrt{K}} \sum_k \Pi_k^2 \rightarrow C_S$. By applying the Lindeberg CLT due to bounded moments,

$$\frac{1}{\sqrt{K}} \begin{pmatrix} \sum_k \Pi_k z_{1k} \\ \sum_k \Pi_{Yk} z_{1k} \\ \sum_k \Pi_{Yk} z_{2k} \\ \sum_k \Pi_k z_{2k} \\ \sum_k z_{1k} z_{2k} \\ \sum_k z_{2k}^2 \\ \sum_k z_{1k}^2 \end{pmatrix} \stackrel{a}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sqrt{K} \omega_{\zeta\eta} \\ \sqrt{K} \omega_{\eta\eta} \\ \sqrt{K} \omega_{\zeta\zeta} \end{pmatrix}, V \right),$$

where V is some variance matrix. By assumption, $\frac{1}{\sqrt{K}} \sum_k \Pi_{Yk} \Pi_k \rightarrow C_Y$ and $\frac{1}{\sqrt{K}} \sum_k \Pi_{Yk}^2 \rightarrow C_{YY}$, so

$$\begin{aligned} \frac{1}{\sqrt{K}} \begin{pmatrix} s_1' s_1 \\ s_1' s_2 \\ s_2' s_2 \end{pmatrix} &= \frac{1}{\sqrt{K}} \begin{pmatrix} \sum_k \Pi_{Yk}^2 + 2 \sum_k \Pi_{Yk} z_{1k} + \sum_k z_{1k}^2 \\ \sum_k \Pi_{Yk} \Pi_k + \sum_k \Pi_{Yk} z_{2k} + \sum_k \Pi_k z_{1k} + \sum_k z_{1k} z_{2k} \\ \sum_k \Pi_k^2 + 2 \sum_k \Pi_k z_{2k} + \sum_k z_{2k}^2 \end{pmatrix} \\ &\stackrel{a}{\sim} \begin{pmatrix} C_{YY} \\ C_Y \\ C \end{pmatrix} + A \frac{1}{\sqrt{K}} \begin{pmatrix} \sum_k \Pi_k z_{1k} \\ \sum_k \Pi_{Yk} z_{1k} \\ \sum_k \Pi_{Yk} z_{2k} \\ \sum_k \Pi_k z_{2k} \\ \sum_k z_{1k} z_{2k} \\ \sum_k z_{2k}^2 \\ \sum_k z_{1k}^2 \end{pmatrix}, \text{ where} \\ A &= \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 0 \end{pmatrix}. \end{aligned}$$

This means:

$$\frac{1}{\sqrt{K}} \begin{pmatrix} s_1' s_1 \\ s_1' s_2 \\ s_2' s_2 \end{pmatrix} \stackrel{a}{\sim} N \left(\begin{pmatrix} C_{YY} + \sqrt{K} \omega_{\zeta\zeta} \\ C_Y + \sqrt{K} \omega_{\zeta\eta} \\ C + \sqrt{K} \omega_{\eta\eta} \end{pmatrix}, A V A' \right).$$

Let $\Sigma = AVA'$ to obtain the result as stated. To derive Σ explicitly, I derive V by applying the Isserlis' Theorem. As a special case of the Isserlis' Theorem for X 's that are multivariate normal and mean zero,

$$E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3].$$

Another corollary is that if n is odd, then there is no such pairing, so the moment is always zero. Hence,

$$\begin{aligned} E[z_{1k}^2 z_{2k}^2] &= E[z_{1k}^2]E[z_{2k}^2] + 2E[z_{1k}z_{2k}]E[z_{1k}z_{2k}] = \omega_{\zeta\zeta}\omega_{\eta\eta} + 2\omega_{\zeta\eta}^2, \text{ and} \\ \text{Var}(z_{1k}z_{2k}) &= \omega_{\zeta\zeta}\omega_{\eta\eta} + \omega_{\zeta\eta}^2. \end{aligned}$$

Similarly,

$$\begin{aligned} \text{Var}(z_{2k}^2) &= E[z_{2k}^4] - \omega_{\eta\eta}^2 = 3\omega_{\eta\eta}^2 - \omega_{\eta\eta}^2 = 2\omega_{\eta\eta}^2, \\ \text{Cov}(z_{1k}, z_{1k}z_{2k}) &= E[z_{1k}^2 z_{2k}] - E[z_{1k}]E[z_{1k}z_{2k}] = 0, \\ \text{Cov}(z_{1k}^2, z_{1k}z_{2k}) &= E[z_{1k}^3 z_{2k}] - E[z_{1k}^2]E[z_{1k}z_{2k}] \\ &= 3\omega_{\zeta\eta}\omega_{\zeta\zeta} - \omega_{\zeta\zeta}\omega_{\zeta\eta} = 2\omega_{\zeta\eta}\omega_{\zeta\zeta}, \\ \text{Cov}(z_{1k}^2, z_{2k}^2) &= E[z_{1k}^2 z_{2k}^2] - \omega_{\zeta\zeta}\omega_{\eta\eta} = 2\omega_{\zeta\eta}^2, \text{ and} \end{aligned}$$

$$V = \begin{bmatrix} \frac{1}{K} \sum_k \Pi_k^2 \omega_{\zeta\zeta} & \frac{1}{K} \sum_k \Pi_k \Pi_{Yk} \omega_{\zeta\zeta} & \frac{1}{K} \sum_k \Pi_k \Pi_{Yk} \omega_{\zeta\eta} & \frac{1}{K} \sum_k \Pi_k^2 \omega_{\zeta\eta} & 0 & 0 & 0 \\ \cdot & \frac{1}{K} \sum_k \Pi_{Yk}^2 \omega_{\zeta\zeta} & \frac{1}{K} \sum_k \Pi_{Yk}^2 \omega_{\zeta\eta} & \frac{1}{K} \sum_k \Pi_k \Pi_{Yk} \omega_{\zeta\eta} & 0 & 0 & 0 \\ \cdot & \cdot & \frac{1}{K} \sum_k \Pi_{Yk}^2 \omega_{\eta\eta} & \frac{1}{K} \sum_k \Pi_k \Pi_{Yk} \omega_{\eta\eta} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \frac{1}{K} \sum_k \Pi_k^2 \omega_{\eta\eta} & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \omega_{\zeta\zeta}\omega_{\eta\eta} + \omega_{\zeta\eta}^2 & 2\omega_{\zeta\eta}\omega_{\eta\eta} & 2\omega_{\zeta\eta}\omega_{\zeta\zeta} \\ \cdot & \cdot & \cdot & \cdot & \cdot & 2\omega_{\eta\eta}^2 & 2\omega_{\zeta\eta}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 2\omega_{\zeta\zeta}^2 \end{bmatrix}.$$

If $\frac{1}{K} \sum_k \Pi_k^2 \rightarrow 0$, $\frac{1}{K} \sum_k \Pi_k \Pi_{Yk} \rightarrow 0$, $\frac{1}{K} \sum_k \Pi_{Yk}^2 \rightarrow 0$ under weak identification, then we obtain the Σ expression stated in the proposition. \square

D.2 Proofs for Section 4.2

Proof of Lemma 2. I work out the μ 's first. Using the judge structure, $\sum_i M_{ii}^2 = \sum_k \frac{(c-1)^2}{c}$, $\sum_i \sum_{j \neq i} P_{ij} = \sum_k \frac{c-1}{c}$. We have also chosen $\pi_k, \sigma_{\xi vk}$ such that $\sum_k \pi_k = 0$, $\sum_k \sigma_{\xi vk} = 0$, $\sum_k \pi_k \sigma_{\xi vk} = 0$. Then, we get the result for means:

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{K}} \sum_k (c-1) (\pi_k^2 \beta^2 + 2\pi_k \beta \sigma_{\xi vk} + \sigma_{\xi vk}^2) \\ \frac{1}{\sqrt{K}} \sum_k (c-1) (\pi_k^2 \beta + \pi_k \sigma_{\xi vk}) \\ \frac{1}{\sqrt{K}} \sum_k (c-1) \pi_k^2 \end{pmatrix} = \begin{pmatrix} \sqrt{K} (c-1) (s^2 \beta^2 + h^2) \\ \sqrt{K} (c-1) s^2 \beta \\ \sqrt{K} (c-1) s^2 \end{pmatrix}.$$

Using a derivation similar to that of the lemma for V_{LM} expression,

$$\begin{aligned} K\sigma_{22} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} (G_{ji} G_{ki} E[\zeta_i^2] R_j R_k + 2G_{ij} G_{ki} E[\eta_i \zeta_i] R_{Yj} R_k + G_{ij} G_{ik} E[\eta_i^2] R_{Yj} R_{Yk}) \\ &\quad + \sum_i \sum_{j \neq i} (G_{ij}^2 E[\eta_i^2] E[\zeta_j^2] + G_{ij} G_{ji} E[\eta_i \zeta_i] E[\eta_j \zeta_j]); \\ K\sigma_{11} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\zeta_i^2] R_{Yj} R_{Yk} (G_{ji} G_{ki} + 2G_{ij} G_{ki} + G_{ij} G_{ik}) + \sum_i \sum_{j \neq i} E[\zeta_i^2] E[\zeta_j^2] (G_{ij}^2 + G_{ij} G_{ji}); \\ K\sigma_{33} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\eta_i^2] R_j R_k (G_{ji} G_{ki} + 2G_{ij} G_{ki} + G_{ij} G_{ik}) + \sum_i \sum_{j \neq i} E[\eta_i^2] E[\eta_j^2] (G_{ij}^2 + G_{ij} G_{ji}); \end{aligned}$$

$$\begin{aligned}
K\sigma_{12} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} (G_{ji}G_{ki}E[\zeta_i^2] R_j R_{Yk} + 2G_{ij}G_{ki}E[\zeta_i^2] R_{Yj}R_k + G_{ij}G_{ik}E[\eta_i\zeta_i] R_{Yj}R_{Yk}) \\
&\quad + \sum_i \sum_{j \neq i} E[\eta_i\zeta_i] E[\zeta_j^2] (G_{ij}^2 + G_{ij}G_{ji}); \\
K\sigma_{23} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} (G_{ji}G_{ki}E[\eta_i^2] R_{Yj}R_k + 2G_{ij}G_{ki}E[\eta_i^2] R_j R_{Yk} + G_{ij}G_{ik}E[\eta_i\zeta_i] R_j R_k) \\
&\quad + \sum_i \sum_{j \neq i} E[\eta_i\zeta_i] E[\eta_j^2] (G_{ij}^2 + G_{ij}G_{ji}); \text{ and} \\
K\sigma_{13} &= \sum_i \sum_{j \neq i} \sum_{k \neq i} E[\eta_i\zeta_i] R_{Yj}R_k (G_{ji}G_{ki} + 2G_{ij}G_{ki} + G_{ij}G_{ik}) + \sum_i \sum_{j \neq i} E[\eta_i\zeta_i] E[\eta_j\zeta_j] (G_{ij}^2 + G_{ij}G_{ji}).
\end{aligned}$$

The equalities hold regardless of whether identification is strong or weak and whether heterogeneity converges or not. Without covariates, $G = P$ is symmetric and the above expressions simplify. For instance,

$$K\sigma_{22} = \sum_k \frac{(c-1)^2}{c} (\omega_{\zeta\zeta k} \pi_k^2 + 2\omega_{\zeta\eta k} \pi_k \pi_{Yk} + \omega_{\eta\eta k} \pi_{Yk}^2) + \sum_k \frac{c-1}{c} (\omega_{\eta\eta k} \omega_{\zeta\zeta k} + \omega_{\zeta\eta k}^2).$$

Evaluate the terms in the expression. For higher moments of π_k , $\sum_k \pi_k^2 = Ks^2$, $\sum_k \pi_k^3 = 0$, $\sum_k \pi_k^4 = Ks^4$. Similarly, $\sum_k \pi_k^3 \sigma_{\xi v} = 0$. Treating the heterogeneity in the same way, $\sum_k \sigma_{\xi v}^2 = Kh^2$. Then,

$$\begin{aligned}
\sum_k \omega_{\zeta\zeta k} \pi_k^2 &= \sum_k (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi v k} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} - \sigma_{\xi v k}^2 + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\xi v k}^2 + 2\sigma_{\varepsilon v} \beta) \pi_k^2 \\
&= s^2 K (s^2 \sigma_{\xi\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v} \beta); \text{ and} \\
\sum_k \omega_{\zeta\eta k} \pi_k \pi_{Yk} &= \sum_k (\pi_k \sigma_{\xi v k} + \sigma_{vv} \beta + \sigma_{\varepsilon v}) \pi_k (\pi_k \beta + \sigma_{\xi v k}) \\
&= \sum_k (\sigma_{vv} \beta^2 \pi_k^2 + \sigma_{\varepsilon v} \pi_k^2 \beta + \pi_k^2 \sigma_{\xi v k}^2) = s^2 K (\sigma_{vv} \beta^2 + \sigma_{\varepsilon v} \beta + h^2).
\end{aligned}$$

Now, for the P_{ij}^2 part,

$$\begin{aligned}
\sum_k \omega_{\eta\eta k} \omega_{\zeta\zeta k} &= \sum_k \sigma_{vv} (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi v k} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} - \sigma_{\xi v k}^2 + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\xi v k}^2 + 2\sigma_{\varepsilon v} \beta) \\
&= \sum_k \sigma_{vv} (\pi_k^2 \sigma_{\xi\xi} + \sigma_{\varepsilon\varepsilon} - \sigma_{\xi v k}^2 + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\xi v k}^2 + 2\sigma_{\varepsilon v} \beta) \\
&= K \sigma_{vv} (s^2 \sigma_{\xi\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v} \beta); \text{ and} \\
\sum_k \omega_{\zeta\eta k}^2 &= \sum_k (\pi_k \sigma_{\xi v k} \pi_k \sigma_{\xi v k} + \sigma_{vv} \beta \pi_k \sigma_{\xi v k} + \sigma_{\varepsilon v} \pi_k \sigma_{\xi v k} + \pi_k \sigma_{\xi v k} \sigma_{vv} \beta + \sigma_{vv} \beta \sigma_{vv} \beta + \sigma_{\varepsilon v} \sigma_{vv} \beta) \\
&\quad + \sum_k (\pi_k \sigma_{\xi v k} \sigma_{\varepsilon v} + \sigma_{vv} \beta \sigma_{\varepsilon v} + \sigma_{\varepsilon v}^2) \\
&= \sum_k (\pi_k^2 \sigma_{\xi v k}^2 + \sigma_{vv}^2 \beta^2 + \sigma_{\varepsilon v} \sigma_{vv} \beta + \sigma_{vv} \beta \sigma_{\varepsilon v} + \sigma_{\varepsilon v}^2) = K (s^2 h^2 + (\sigma_{vv} \beta + \sigma_{\varepsilon v})^2).
\end{aligned}$$

Combine the expressions for σ_{22} and impose asymptotics where $s \rightarrow 0$ and $h \rightarrow 0$:

$$\begin{aligned}
\sigma_{22} &= \frac{1}{K} \sum_k \frac{(c-1)^2}{c} h^2 + \frac{1}{K} \sum_k \frac{c-1}{c} (\sigma_{vv} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v} \beta) + (\sigma_{vv} \beta + \sigma_{\varepsilon v})^2) + o(1) \\
&= \frac{c-1}{c} (\sigma_{vv} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta) + (\sigma_{vv} \beta + \sigma_{\varepsilon v})^2) + o(1).
\end{aligned}$$

Next, evaluate a few more sums that feature in the other σ expressions:

$$\begin{aligned}
\sum_k \omega_{\zeta\zeta} \pi_{Yk}^2 &= \sum_k (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi vk} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta) (\pi_k^2 \beta^2 + 2\pi_k \sigma_{\xi vk} + \sigma_{\xi v}^2) \\
\frac{1}{K} \sum_k \omega_{\zeta\zeta} \pi_{Yk}^2 &= \frac{1}{K} \sum_k \sigma_{\xi v}^2 (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi vk} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta) \\
&= h^2 (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v} \beta) = o(1); \\
\frac{1}{K} \sum_k \omega_{\zeta\zeta}^2 &= \frac{1}{K} \sum_k (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi vk} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta)^2 \\
&= \frac{1}{K} \sum_k (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta)^2 = (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta)^2; \\
\frac{1}{K} \sum_k \omega_{\zeta\eta} \pi_{Yk}^2 &= \frac{1}{K} \sum_k (\pi_k \sigma_{\xi vk} + \sigma_{vv} \beta + \sigma_{\varepsilon v}) (\pi_k^2 \beta^2 + 2\pi_k \sigma_{\xi vk} + \sigma_{\xi v}^2) \\
&= h^2 (\sigma_{vv} \beta + \sigma_{\varepsilon v}) = o(1); \text{ and} \\
\frac{1}{K} \sum_k \omega_{\zeta\eta} \omega_{\zeta\zeta} &= \frac{1}{K} \sum_k (\pi_k \sigma_{\xi vk} + \sigma_{vv} \beta + \sigma_{\varepsilon v}) (\pi_k^2 \sigma_{\xi\xi} + 2\pi_k \beta \sigma_{\xi vk} + 2\pi_k \sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta) \\
&= \frac{1}{K} \sum_k (\sigma_{vv} \beta + \sigma_{\varepsilon v}) (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + \sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v} \beta) \\
&= (\sigma_{vv} \beta + \sigma_{\varepsilon v}) (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta) + o(1).
\end{aligned}$$

Using these results,

$$\begin{aligned}
\sigma_{22} &= \frac{c-1}{c} (\sigma_{vv} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta) + (\sigma_{vv} \beta + \sigma_{\varepsilon v})^2) + o(1); \\
\sigma_{11} &= 2 \frac{c-1}{c} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta)^2 + o(1); \\
\sigma_{33} &= 2 \frac{c-1}{c} \sigma_{vv}^2 + o(1); \\
\sigma_{12} &= 2 \frac{c-1}{c} (\sigma_{vv} \beta + \sigma_{\varepsilon v}) (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta) + o(1); \\
\sigma_{23} &= 2 \frac{c-1}{c} \sigma_{vv} (\sigma_{vv} \beta + \sigma_{\varepsilon v}) + o(1); \text{ and} \\
\sigma_{13} &= 2 \frac{c-1}{c} (\sigma_{vv} \beta + \sigma_{\varepsilon v})^2 + o(1).
\end{aligned}$$

Hence, $\sigma_{13} = \sigma_{23}^2 / \sigma_{33} + o(1)$ is immediate. Further, for σ_{12} ,

$$\begin{aligned}
2 \frac{\sigma_{23}}{\sigma_{33}} \left(\sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right) &= 2 \frac{\sigma_{vv} \beta + \sigma_{\varepsilon v}}{\sigma_{vv}} \left(\sigma_{22} - \frac{(2 \frac{c-1}{c} \sigma_{vv} (\sigma_{vv} \beta + \sigma_{\varepsilon v}))^2}{2 \times 2 \frac{c-1}{c} \sigma_{vv}^2} \right) + o(1) \\
&= 2 \frac{c-1}{c} (\sigma_{vv} \beta + \sigma_{\varepsilon v}) (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta) + o(1) = \sigma_{12} + o(1).
\end{aligned}$$

Finally, the σ_{11} can be obtained:

$$\begin{aligned}
\frac{4}{\sigma_{33}} \left(\sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right)^2 &= \frac{2}{\frac{c-1}{c} \sigma_{vv}^2} \left(\frac{c-1}{c} (\sigma_{vv} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v} \beta)) \right)^2 + o(1) \\
&= 2 \frac{c-1}{c} (\sigma_{\varepsilon\varepsilon} + \sigma_{vv} \beta^2 + \sigma_{vv} \sigma_{\xi\xi} + 2\sigma_{\varepsilon v} \beta)^2 + o(1) = \sigma_{11} + o(1).
\end{aligned}$$

□

Proof of Proposition 2. The first two are straightforward: $C_S = \mu_3/(c-1)$ and $\beta = \mu_2/\mu_3$ imply $\mu_3 = (c-1)C_S$ and $\mu_2 = (c-1)C_S\beta$. For μ_1 , observe that:

$$\begin{aligned} h &= \sqrt{\frac{1}{\sqrt{K}} \frac{1}{c-1} \left(\mu_1 - \frac{\mu_2^2}{\mu_3} \right)} = \sqrt{\frac{1}{\sqrt{K}} (\mu_1 - C_S\beta^2)}, \text{ and} \\ C_H &= \sqrt{K}h^2 = \mu_1/(c-1) - C_S\beta^2, \text{ so} \\ (c-1)(C_S\beta^2 + C_H) &= (c-1)(C_S\beta^2 + \mu_1/(c-1) - C_S\beta^2) = \mu_1 \end{aligned}$$

as required. Next, since $\sigma_{vv} = \sqrt{\frac{\sigma_{33}c}{2(c-1)}}$, $\sigma_{33} = 2\frac{c-1}{c}\sigma_{vv}^2$ is immediate. Similarly, with $\sigma_{\varepsilon v} = \frac{1}{\sigma_{vv}} \left(\frac{\sigma_{23}c}{2(c-1)} - \sigma_{vv}^2\beta \right)$, $\sigma_{23} = 2\frac{c-1}{c}\sigma_{vv}(\sigma_{vv}\beta + \sigma_{\varepsilon v})$. From these two expressions, we can observe that:

$$(\sigma_{vv}\beta + \sigma_{\varepsilon v})^2 = \frac{c}{2(c-1)} \frac{\sigma_{23}^2}{\sigma_{33}}.$$

To obtain an expression for σ_{22} , rearrange $\sigma_{\varepsilon\varepsilon} = \frac{1}{\sigma_{vv}} \frac{c}{c-1} \left(\sigma_{22} - \frac{\sigma_{23}^2}{\sigma_{33}} \right) + \frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}} \geq 0$:

$$\begin{aligned} \sigma_{22} &= \frac{\sigma_{23}^2}{\sigma_{33}} + \frac{c-1}{c} (\sigma_{\varepsilon\varepsilon}\sigma_{vv} - \sigma_{\varepsilon v}^2) \\ &= \frac{c-1}{c} \left(\sigma_{vv}(\sigma_{\varepsilon\varepsilon} + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\varepsilon v}\beta) + (\sigma_{vv}\beta + \sigma_{\varepsilon v})^2 \right) + o(1), \end{aligned}$$

where the final step uses $\sigma_{\xi\xi} = h/\sigma_{vv}$. This expression for σ_{22} is of the form required in Lemma 2. Then,

$$\begin{aligned} \det(\Sigma_{SF}) &= \sigma_{\varepsilon\varepsilon}\sigma_{\xi\xi}\sigma_{vv} - \sigma_{\varepsilon\varepsilon}h^2 - \sigma_{\xi\xi}^2\sigma_{vv} + 2\sigma_{\varepsilon\xi}\sigma_{\varepsilon v}h - \sigma_{\xi\xi}\sigma_{\varepsilon v}^2 \\ &= \sigma_{\varepsilon\varepsilon}\sigma_{\xi\xi}\sigma_{vv} - \sigma_{\varepsilon\varepsilon}h^2 - \sigma_{\xi\xi}\sigma_{\varepsilon v}^2 = \sigma_{\varepsilon\varepsilon}h - \sigma_{\varepsilon\varepsilon}h^2 - h\frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}}; \text{ and} \\ \det(\Sigma_{SF})/h &= \sigma_{\varepsilon\varepsilon} - \frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}} - \sigma_{\varepsilon\varepsilon}h = \sigma_{\varepsilon\varepsilon} - \frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}} + o(1). \end{aligned}$$

An analogous argument holds for $\sigma_{\xi vk} = -h$. From the σ_{22} equation, $\sigma_{\varepsilon\varepsilon} - \frac{\sigma_{\varepsilon v}^2}{\sigma_{vv}} = \frac{c}{c-1} \left(\sigma_{22} - \frac{\sigma_{23}^2}{\sigma_{33}} \right) \geq 0$, which delivers the result that $\det(\Sigma_{SF})/h \rightarrow C_{det} \geq 0$. \square

D.3 Proofs for Section 4.3

Proof of Proposition 3. Fix any alternative $(\pi^A, \pi_Y^A) \in \mathcal{S}$ with corresponding $(\mu_1^A, \mu_2^A, \mu_3^A)$. Due to the restriction in \mathcal{S} ,

$$\begin{pmatrix} \mu_1^H \\ \mu_2^H \\ \mu_3^H \end{pmatrix} = \begin{pmatrix} \mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A \\ 0 \\ \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A \end{pmatrix} = L \begin{pmatrix} \mu_1^H \\ \mu_2^H \\ \mu_3^H \end{pmatrix}$$

is in the null space. Construct Neyman-Pearson test for μ^H vs μ^A . The Neyman-Pearson test rejects for large values of:

$$\log \frac{dN(\mu^A, \Sigma)}{dN(\mu^H, \Sigma)} = \frac{\mu_2^A}{\sigma_{22}} X_2 - \frac{1}{2} \frac{(\mu_2^A)^2}{\sigma_{22}}.$$

Hence, the most powerful test rejects large values of X_2 , which is what LM does. By [Lehmann and Romano \(2005\)](#) Theorem 3.8.1(i), since LM is valid for any distribution in the null space (by Theorem 1) and it is most powerful for some distribution in the null space, LM is most powerful for testing the composite null against the given alternative (π^A, π_Y^A) . \square

Proof of Proposition 4. Let $\mu \in \mathcal{M} = \{\mu : \mu_1 > 0, \mu_3 > 0, \mu_2^2 < \mu_1\mu_3\}$. I first show that \mathcal{M} is con-

vex. For $\lambda \in (0, 1)$, it suffices to show, for μ_a and μ_b that satisfy $\mu_{2a}^2 < \mu_{1a}\mu_{3a}$ and $\mu_{2b}^2 < \mu_{1b}\mu_{3b}$, that $(\lambda\mu_{2a} + (1-\lambda)\mu_{2b})^2 < (\lambda\mu_{1a} + (1-\lambda)\mu_{1b})(\lambda\mu_{3a} + (1-\lambda)\mu_{3b})$. This set is intersected with the set that satisfies $\mu_1 > 0$ and $\mu_3 > 0$, which is clearly convex. The following is negative:

$$\begin{aligned}
& (\lambda\mu_{2a} + (1-\lambda)\mu_{2b})^2 - (\lambda\mu_{1a} + (1-\lambda)\mu_{1b})(\lambda\mu_{3a} + (1-\lambda)\mu_{3b}) \\
&= \lambda^2\mu_{2a}^2 + (1-\lambda)^2\mu_{2b}^2 + 2\lambda(1-\lambda)\mu_{2a}\mu_{2b} - \lambda^2\mu_{1a}\mu_{3a} - (1-\lambda)^2\mu_{1b}\mu_{3b} - \lambda(1-\lambda)(\mu_{1b}\mu_{3a} + \mu_{1a}\mu_{3b}) \\
&= \lambda^2(\mu_{2a}^2 - \mu_{1a}\mu_{3a}) + (1-\lambda)^2(\mu_{2b}^2 - \mu_{1b}\mu_{3b}) + \lambda(1-\lambda)(2\mu_{2a}\mu_{2b} - \mu_{1b}\mu_{3a} - \mu_{1a}\mu_{3b}) \\
&< \lambda(1-\lambda)(2\sqrt{\mu_{1a}\mu_{1b}\mu_{1b}\mu_{3b}} - \mu_{1b}\mu_{3a} - \mu_{1a}\mu_{3b}) \\
&< -\lambda(1-\lambda)(\sqrt{\mu_{1b}\mu_{3a}} - \sqrt{\mu_{1a}\mu_{3b}})^2 \leq 0.
\end{aligned}$$

The first inequality occurs from applying $\mu_{2a}^2 < \mu_{1a}\mu_{3a}$ and $\mu_{2b}^2 < \mu_{1b}\mu_{3b}$, so \mathcal{M} is convex. Let $m \sim N(\mu, \Sigma)$ denote a statistic drawn from the asymptotic distribution, with m_i being a component of the vector m , so that m_2 is the LM statistic. Using the linear transformation from [Lehmann and Romano \(2005\)](#) Example 3.9.2 Case 3, we can transform the statistics and parameter such that m_2 is orthogonal to all other components. In particular, let L be a 3 by 3 matrix that is a function of Σ . Then, $Lm \sim N(L\mu, I_3)$ with $[Lm]_2 = m_2/\sqrt{\text{Var}(m_2)}$. The parameter space of $L\mu \in \mathcal{L}$ is also convex because L is a linear transformation: take any $\mu_a, \mu_b \in \mathcal{M}$, then observe that $\lambda L\mu_a + (1-\lambda)L\mu_b = L(\lambda\mu_a + (1-\lambda)\mu_b)$. Since \mathcal{M} is convex, and every element in \mathcal{M} is linearly transformed into the space on \mathcal{L} , we have $\lambda\mu_a + (1-\lambda)\mu_b \in \mathcal{M}$ and hence $L(\lambda\mu_a + (1-\lambda)\mu_b) \in \mathcal{L}$. Since Lm is normally distributed and \mathcal{L} is convex with rank 3, the problem is in the exponential class, using the definition from [Lehmann and Romano \(2005\)](#) Section 4.4. Since the joint distribution is in the exponential class and the restriction to the interior ensures that there are points in the parameter space that are above and below the null, the uniformly most powerful unbiased test follows the form of [Lehmann and Romano \(2005\)](#) Theorem 4.4.1(iv). To calculate the critical values of the [Lehmann and Romano \(2005\)](#) Theorem 4.4.1(iv) result, observe that $[Lm]_2$ is orthogonal to $[Lm]_1$ and $[Lm]_3$, so the distribution of $[Lm]_2$ conditional on $[Lm]_1$ and $[Lm]_3$ is standard normal. Since $[Lm]_2$ is standard normal, it is symmetric around 0 under the null, so the solution to the critical value is ± 1.96 for a 5% test, due to simplification in [Lehmann and Romano \(2005\)](#) Section 4.2. The resulting test is hence identical to the two-sided LM test. \square

E Proofs for Section 5

Proof of Lemma 3. The A expressions can be written as:

$$\begin{aligned}
A_1 &= \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} G_{ij} X_j G_{ik} X_k (Y_i Y_l - X_i Y_l \beta_0 - Y_i X_l \beta_0 + X_i X_l \beta_0^2); \\
A_2 &= \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} G_{ij} X_j G_{ki} X_l (Y_i Y_k - X_i Y_k \beta_0 - Y_i X_k \beta_0 + X_i X_k \beta_0^2); \\
A_3 &= \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} X_l G_{ji} G_{ki} X_i (Y_j Y_k - X_j Y_k \beta_0 - Y_j X_k \beta_0 + X_j X_k \beta_0^2); \\
A_4 &= \sum_i \sum_{j \neq i} \sum_{k \neq j} \sum_{l \neq i, k} \check{M}_{jl, -ijk} \check{M}_{ik, -ij} G_{ji}^2 X_i X_k (Y_j Y_l - X_j Y_l \beta_0 - Y_j X_l \beta_0 + X_j X_l \beta_0^2); \text{ and} \\
A_5 &= \sum_i \sum_{j \neq i} \sum_{k \neq j} \sum_{l \neq i, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} G_{ij} G_{ji} X_k X_l (Y_i Y_j - X_i Y_j \beta_0 - Y_i X_j \beta_0 + X_i X_j \beta_0^2).
\end{aligned}$$

Since these terms have a quadratic form, the variance estimator is also quadratic in β_0^2 , i.e.,

$$\hat{V}_{LM} = C_0 + C_1 \beta_0 + C_2 \beta_0^2,$$

where the C 's can be worked out by collecting the expressions above. For instance,

$$\begin{aligned} C_0 = & \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} G_{ij} X_j G_{ik} X_k Y_i Y_l + 2 \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} G_{ij} X_j G_{ki} X_l Y_i Y_k \\ & + \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \check{M}_{il, -ijk} X_l G_{ji} G_{ki} X_i Y_j Y_k \\ & - \sum_i \sum_{j \neq i} \sum_{k \neq j} \sum_{l \neq i, k} \check{M}_{jl, -ijk} \check{M}_{ik, -ij} G_{ji}^2 X_i X_k Y_j Y_l - \sum_i \sum_{j \neq i} \sum_{k \neq j} \sum_{l \neq i, k} \check{M}_{ik, -ij} \check{M}_{jl, -ijk} G_{ij} G_{ji} X_k X_l Y_i Y_j \end{aligned}$$

C_1 and C_2 are analogous by collecting the coefficients on β_0, β_0^2 from expressions A_1 to A_5 . The test does not reject:

$$\frac{(P_{XY} - P_{XX}\beta_0)^2}{C_0 + C_1\beta_0 + C_2\beta_0^2} \leq q \Leftrightarrow (P_{XX}^2 - qC_2) \beta_0^2 - (2P_{XY}P_{XX} + qC_1) \beta_0 + (P_{XY}^2 - qC_0) \leq 0.$$

Solutions exist when:

$$D := (2P_{XY}P_{XX} + qC_1)^2 - 4(P_{XX}^2 - qC_2)(P_{XY}^2 - qC_0) \geq 0.$$

The rest of the lemma are immediate from properties of solving quadratic inequalities. \square

F Derivations for Section 6

Derivation for binary setup without covariates.

The reduced-form residuals are given by:

$$\eta_i \mid v_i = \begin{cases} 1 - \pi_k & \text{if } v_i \leq \pi_k \\ -\pi_k & \text{if } v_i > \pi_k \end{cases}, \text{ and } \zeta_i = \pi_{k(i)}\beta_i - \pi_{Yk(i)} + \eta_i\beta_i + \varepsilon_i.$$

Imposing $E[\zeta_i] = 0$, $\pi_{Yk(i)} = \pi_{k(i)}\beta + E[\eta_i\beta_i]$, where $E[\eta_i\beta_i] = -(1-s)(2p-1)\sigma_{\xi vk}$. Hence,

$$\pi_{Yk} = \pi_k\beta - (1-s)(2p-1)\sigma_{\xi vk}.$$

Following the same argument as Section 2, due to the judge setup, the estimand is:

$$\frac{\sum_k \pi_{Yk} \pi_k}{\sum_k \pi_k^2} = \frac{\sum_k (\pi_k\beta - (1-s)(2p-1)\sigma_{\xi vk}) \pi_k}{\sum_k \pi_k^2} = \beta$$

because $\sum_k \sigma_{\xi vk} \pi_k = 0$ by construction.

Derivation for binary setup with covariates.

Consider the structural model:

$$\begin{aligned} Y_i(x) &= x(\beta + \xi_i) + w'\gamma + \varepsilon_i, \text{ and} \\ X_i(z) &= I\{z'\pi + w'\gamma - v_i \geq 0\}. \end{aligned}$$

Let \mathcal{N}_t denote the set of observations in state t . Then, using the G that corresponds to UJIVE,

$$\begin{aligned} \sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} G_{ij} R_{Yi} R_j &= \sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} G_{ij} (\pi_{Yk(i)} + \gamma_{t(i)}) (\pi_{k(j)} + \gamma_{t(j)}) \\ &= \sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} G_{ij} (\pi_{Yk(i)} \pi_{k(j)} + \gamma_{t(i)} \pi_{k(j)} + \pi_{Yk(i)} \gamma_{t(j)} + \gamma_{t(i)} \gamma_{t(j)}) \\ &= \frac{1}{1-1/5} \sum_{k \in \{0, t\}} 5 \times 4 \times \frac{1}{5} (\pi_{Yk} \pi_k + \gamma_t \pi_k + \pi_{Yk} \gamma_t + \gamma_t^2) \end{aligned}$$

$$\begin{aligned}
& -\frac{1}{1-1/10} \sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} \frac{1}{10} (\pi_{Yk(i)} \pi_{k(j)} + \gamma_t \pi_{k(j)} + \pi_{Yk(i)} \gamma_t + \gamma_t^2) \\
& = \sum_{k \in \{0, t\}} 5 (\pi_{Yk} \pi_k + \gamma_t \pi_k + \pi_{Yk} \gamma_t + \gamma_t^2) - \frac{1}{9} \sum_{k \in \{0, t\}} 5 \times 4 (\pi_{Yk} \pi_k + \gamma_{Yt} \pi_k + \pi_{Yk} \gamma_{Xt} + \gamma_t^2) \\
& \quad - \frac{1}{9} 5 \times 5 (\pi_{Yt} \pi_0 + \gamma_t \pi_0 + \pi_{Yt} \gamma_t + \gamma_t^2) - \frac{1}{9} 5 \times 5 (\pi_{Y0} \pi_t + \gamma_t \pi_t + \pi_{Y0} \gamma_t + \gamma_t^2) \\
& = 5 \left(\frac{5}{9} \right) (\pi_{Y0} \pi_0 + \pi_{Yt} \pi_t - \pi_{Yt} \pi_0 - \pi_{Y0} \pi_t).
\end{aligned}$$

Recall that $\pi_{Yk} = \pi_k \beta - (1-s)(2p-1)\sigma_{\xi vk}$. Impose $\sigma_{\xi v0} = 0$ for all k , so that $\pi_{Y0} = 0$, because I have already set $\pi_0 = 0$. Then,

$$\pi_{Yk} = \pi_k \beta - (1-s)(2p-1)\sigma_{\xi vk}.$$

Using the result directly,

$$\sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} G_{ij} R_{Yi} R_j = 5 \left(\frac{5}{9} \right) (\pi_{Y0} \pi_0 + \pi_{Yt} \pi_t - \pi_{Yt} \pi_0 - \pi_{Y0} \pi_t) = \frac{25}{9} \pi_{Yt} \pi_t.$$

Analogously, $\sum_{i \in \mathcal{N}_t} \sum_{j \in \mathcal{N}_t \setminus i} G_{ij} R_i R_j = \frac{25}{9} \pi_t^2$. Hence, as long as $\sum_t \sigma_{\xi vt} \pi_t = 0$, which is the case for the construction in the main text, we still recover β as our estimand:

$$\begin{aligned}
\frac{\sum_i \sum_{j \neq i} G_{ij} R_{Yi} R_j}{\sum_i \sum_{j \neq i} G_{ij} R_i R_j} &= \frac{\sum_t \pi_{Yt} \pi_t}{\sum_t \pi_t^2} = \frac{\sum_t (\pi_t \beta - (1-s)(2p-1)\sigma_{\xi vt}) \pi_t}{\sum_t \pi_t^2} \\
&= \beta - \frac{\sum_t (1-s)(2p-1)\sigma_{\xi vt} \pi_t}{\sum_t \pi_t^2} = \beta.
\end{aligned}$$

This happens in our construction regardless of what happens to γ_t .

G Proofs for Appendix A

Proof of Lemma 4. Suppose not. Then, for some real β_0 ,

$$E[T_{ee}] = \sum_i \sum_{j \neq i} P_{ij} R_{\Delta i} R_{\Delta j} = \sum_i \sum_{j \neq i} P_{ij} (R_{Yi} R_{Yj} - R_i R_{Yj} \beta_0 - R_{Yi} R_j \beta_0 + R_i R_j \beta_0^2) = 0.$$

Solving for β_0 ,

$$\beta_0 = \frac{2 \sum_i \sum_{j \neq i} P_{ij} R_i R_{Yj} \pm \sqrt{4 \left(\sum_i \sum_{j \neq i} P_{ij} R_i R_{Yj} \right)^2 - 4 \left(\sum_i \sum_{j \neq i} P_{ij} R_i R_j \right) \left(\sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_{Yj} \right)}}{2 \left(\sum_i \sum_{j \neq i} P_{ij} R_i R_j \right)}.$$

In our structural model, $R_i = \pi_{k(i)}$ and $R_{Yi} = \pi_{Yk(i)}$. The term in the square root can be written as:

$$D = 4 \left(\sum_k \pi_k \pi_{Yk} \right)^2 - 4 \left(\sum_k \pi_k^2 \right) \left(\sum_k \pi_{Yk}^2 \right)$$

Since $\pi_{Yk} = \pi_k \beta + \sigma_{\xi vk}$, by imposing $\sum_k \pi_k = 0$,

$$\frac{1}{4} D = \left(\sum_k \pi_k^2 \beta + \sum_k \pi_k \sigma_{\xi vk} \right)^2 - \left(\sum_k \pi_k^2 \right) \left(\sum_k (\pi_k^2 \beta^2 + \sigma_{\xi vk}^2 + 2\pi_k \beta \sigma_{\xi vk}) \right)$$

$$= \left(\sum_k \pi_k^2 \beta \right)^2 - \left(\sum_k \pi_k^2 \right) \left(\sum_k \pi_k^2 \beta^2 + K \sigma_{\xi vk}^2 \right) = - \left(\sum_k \pi_k^2 \right) K \sigma_{\xi vk}^2 \leq 0.$$

Since $\sigma_{\xi vk} \neq 0$ and $\sum_k \pi_k^2 > 0$, there are no real roots of β_0 , a contradiction. \square

Proof of Equation (17).

$$\begin{aligned} E[\hat{\Psi}_{MO}] &= E \left[\sum_i \left(\sum_{j \neq i} P_{ij} (R_j + \eta_j) \right)^2 (R_{\Delta i} + \nu_i)^2 + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i + \eta_i) (R_{\Delta i} + \nu_i) (R_j + \eta_j) (R_{\Delta j} + \nu_j) \right] \\ &= E \left[\sum_i \left(\left(\sum_{j \neq i} P_{ij} R_j \right)^2 + \left(\sum_{j \neq i} P_{ij} \eta_j \right)^2 \right) (R_{\Delta i} + \nu_i)^2 \right] \\ &\quad + E \left[\sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} + \eta_i R_{\Delta i} + R_i \nu_i + \eta_i \nu_i) (R_j R_{\Delta j} + \eta_j R_{\Delta j} + R_j \nu_j + \eta_j \nu_j) \right] \\ &= \sum_i M_{ii}^2 R_i^2 (R_{\Delta i}^2 + E[\nu_i^2]) + \sum_i R_{\Delta i}^2 E \left[\left(\sum_{j \neq i} P_{ij} \eta_j \right)^2 \right] + \sum_i E[\nu_i^2] E \left[\left(\sum_{j \neq i} P_{ij} \eta_j \right)^2 \right] \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} + E[\eta_i \nu_i]) (R_j R_{\Delta j} + E[\eta_j \nu_j]) \\ &= \sum_i M_{ii}^2 R_i^2 (R_{\Delta i}^2 + E[\nu_i^2]) + \sum_i \sum_{j \neq i} P_{ij}^2 E[\eta_j^2 (R_{\Delta i}^2 + \nu_i^2)] + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} + E[\eta_i \nu_i]) (R_j R_{\Delta j} + E[\eta_j \nu_j]) \\ &= \sum_i M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_i M_{ii}^2 R_i^2 E[\nu_i^2] + \sum_i \sum_{j \neq i} P_{ij}^2 E[\nu_i^2] E[\eta_j^2] + \sum_i \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E[\eta_j^2] \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 (R_i R_{\Delta i} R_j R_{\Delta j} + E[\eta_i \nu_i] R_j R_{\Delta j} + R_i R_{\Delta i} E[\eta_j \nu_j] + E[\eta_i \nu_i] E[\eta_j \nu_j]) \end{aligned}$$

\square

As a corollary of Equation (9) and Equation (17),

$$\begin{aligned} Var \left(\sum_i \sum_{j \neq i} P_{ij} e_i X_j \right) - E[\hat{\Psi}_{MO}] &= \sum_i M_{ii}^2 (2 R_i R_{\Delta i} E[\nu_i \eta_i] + E[\eta_i^2] R_{\Delta i}^2 - R_i^2 R_{\Delta i}^2) \\ &\quad - \sum_i \sum_{j \neq i} P_{ij}^2 (R_{\Delta i}^2 E[\eta_j^2] + R_i R_{\Delta i} R_j R_{\Delta j} + E[\eta_i \nu_i] R_j R_{\Delta j} + R_i R_{\Delta i} E[\eta_j \nu_j]), \end{aligned}$$

which reflects the bias of the estimator.