# Inference with Many Weak Instruments and Heterogeneity

Luther Yap*

January 24, 2025

**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. With heterogeneity, the score test is also the uniformly most powerful unbiased test in the asymptotic distribution.

## 1 Introduction

Many empirical studies in economics involve instrumental variable (IV) models with many instruments. A prominent example is the judge design: several studies argue that judges or case workers are as good as randomly assigned and can affect the treatment status, so they are used 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 misdemeanor prosecution (Agan et al., 2023), among others. When the IV is a vector of indicators for judges, the number of instruments can be large relative to the sample size. Another example of many IV is a single instrument interacted with discrete covariates. For instance, when Angrist and Krueger (1991) used the quarter of birth as an instrument to study the returns to education, interacting the quarter of birth with the state of birth can generate 150 instruments.

Recent econometric research also suggests that many instruments should be used. With covariates, Blandhol et al. (2022) show that the standard two-stage least squares (TSLS) estimator puts negative weights on local average treatment effects (LATE), unless there is a parametric model or the regression is fully saturated (i.e., where instruments are fully interacted with covariates). Further, with a saturated regression, several jackknife estimators recover a positively weighted average

*Email: lyap@princeton.edu.

of LATE's. (Evdokimov and Kolesár, 2018; Chao et al., 2023; Boot and Nibbering, 2024) Unless there are a few (or no) discrete covariates, fully interacting the instrument with covariates creates many instruments, further motivating the many IV setting.

Despite the pervasiveness and importance 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 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 IV may estimate different LATE's. It is understood that the standard TSLS estimator for IV is inconsistent and its $t$-statistic test is invalid for inference in the many instrument environment (e.g., Bekker (1994); Bound et al. (1995); Donald and Newey (2001)). While the jackknife IV estimator (JIVE) (e.g., Phillips and Hale (1977); Angrist et al. (1999); Chao et al. (2012)) addresses the estimation problem, its $t$-statistic test 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 are robust to weak IV, but they assume constant treatment effects. A separate literature (Evdokimov and Kolesár, 2018) proposed variance estimators for the JIVE that are robust to heterogeneous treatment effects, but their $t$-statistic test is still not robust to many weak IV. While it is clear that weak IV can lead to substantial distortions in inference (e.g., Dufour (1997); 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) (MS22), which 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) (EK18), which is robust to heterogeneity but not weak instruments, can be severely oversized. Additionally, this section documents how an empirically common practice of constructing a "leniency measure" that combines the many instruments and then using weak IV robust procedures from the just-identified IV literature is invalid.

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 result uses an asymptotic environment that nests the asymptotic environments of EK18 and MS22 in that normality holds if either the number of instruments is large or the instruments are strong. 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. Obtaining a consistent variance estimator is the second technical challenge of the paper, since reduced-form coefficients are not consistently estimable when there are few observations per instrument. Motivated by Anatolyev and Sølvsten (2023) who proposed a method to jointly test the significance of many covariates in OLS, I construct a leave-three-out (L3O) variance estimator for the LM variance and show that it is consistent, even when reduced-form coefficients are not consistently estimable. 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 powerful. In the environment with a fixed reduced-form covariance matrix, I focus on a class of tests that are functions of a natural set of statistics. Then, I show theoretically in the asymptotic distribution 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 for the interior of the alternative space (i.e., when heterogeneity is imposed and constant treatment effects is ruled out). Beyond the scope of the theory, numerical results also suggest that LM is close to a power envelope in an empirical application.

Section 5 shows how the test can be inverted to construct a confidence set that can be expressed in closed form. Simulation results in Section 6 show how the procedure is robust even with a small number of instruments, and it is reasonably powerful even with constant treatment effects. Section 6 also contains two empirical applications that show how being robust to many weak IV and heterogeneity can change conclusions. In the Angrist and Krueger (1991) quarter of birth application, the MS and Matsushita and Otsu (2022) procedures that are robust to many weak IV but not heterogeneity have unbounded confidence sets while L3O has a bounded confidence set. In the Agan et al. (2023) judge application, MS has an empty confidence set, and the length of the L3O confidence interval is more than twice that of EK18 that is not robust to many weak IV.

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 dealing with weak instruments (e.g., Staiger and Stock (1997); Stock and Yogo (2005); Lee et al. (2023)). While recent procedures accommodate both simultaneously (e.g., Crudu et al. (2021); Mikusheva and Sun (2022); Matsushita and Otsu (2022); Yap (2023); Lim et al. (2024)), their focus has been on the linear IV model with constant treatment effects. This paper augments their setup by allowing for heterogeneity in treatment effects, and contributes new results on the limitations of their procedures under heterogeneity. Further, I show how heterogeneity can be understood in a framework analogous to weak instruments.

Second, this paper contributes to the literature on heterogeneous treatment effects (e.g., Kolesár (2013); Evdokimov and Kolesár (2018); Blandhol et al. (2022)). The previous papers exploit consistent estimation of the object of interest to conduct inference. In contrast, this paper uses the (more general) weak IV environment where the object of interest may not be consistently estimated. One contemporaneous paper that allows weak IV and heterogeneity is 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 Sølvsten (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 Sølvsten (2023) that does not require an inversion of an $n$ by $n$ matrix, and whose L3O implementation is fast when using matrix operations.

Fourth, this paper contributes to a literature on optimal tests. While the uniformly most powerful unbiased (UMPU) test for just-identified IV has been established since Moreira (2009b), obtaining a UMPU test in the over-identified IV environment has thus far been more challenging. In the over-identified IV environment with constant treatment effects, several statistics are informative of the object of interest. Consequently, there is a large literature that numerically compares various valid tests and characterizes various forms of optimality (e.g., Moreira (2003); Andrews (2016); Andrews et al. (2019); Van de Sijpe and Windmeijer (2023); Lim et al. (2024)). By imposing heterogeneity in the environment, the problem is (somewhat surprisingly) simplified. Since only one statistic in the asymptotic distribution is directly informative of the object of interest, I can obtain a UMPU result.

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 and empirical applications; 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 potential outcomes model without covariates that exhibits weak instruments and heterogeneity in treatment effects. This model is a special case of the model in Section 3, which presents an inference procedure that is valid for a general model that also accommodates potentially many covariates. A simulation from the model shows how weak instruments and heterogeneity can lead to substantial distortions in inference for procedures recently proposed in the econometric literature. 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 has a rejection rate that is close to the nominal rate.

### 2.1 Setting for Simple Example

The simple example uses the canonical latent variable framework of Heckman and Vytlacil (2005). We are interested in the effect of $X_i \in \{0,1\}$ (e.g., incarceration) on some outcome $Y_i$, for $i = 1, \cdots n$ that indexes individuals. To instrument for $X_i$, we use a vector of judges indicators: $Z_i$ is a $(K+1)$-dimensional vector of indicators for judges, indexed $1, \cdots, K+1$, each with $c = 5$ individual cases, so the vector takes value 1 for the $k$th component when individual $i$ is matched to judge $k$, and 0 elsewhere. Let $Y_i(0)$ and $Y_i(1)$ denote the untreated and treated potential outcomes respectively, and we observe $Y_i = Y_i(X_i)$. The treatment status given some instrument value $z$ is $X_i(z)$, and we observe $X_i(Z_i)$. The model is:

$$X_i(z) = 1\{z'\lambda > v_i\}, \text{ and } Y_i(x) = xf(v_i) + \varepsilon_i, \tag{1}$$

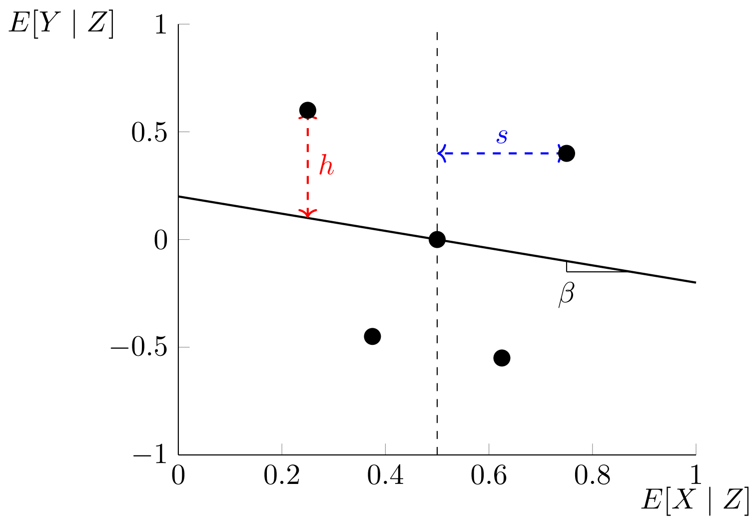
where $1\{\cdot\}$ is an indicator function that takes the value 1 if the argument is true and 0 otherwise. Here, $Z_i'\lambda = \lambda_{k(i)}$, where $k(i)$ is the judge that individual $i$ is matched to. With individual unobservable $v_i \sim U[0,1]$, the probability of treatment (i.e., $X_i = 1$) given judge $k$ is $\lambda_k$. I set $\lambda_k = 1/2$ for the base judge, and evenly split all other $K$ judges to take 4 different values of $\lambda_k$ denoted in Table 1. Potential outcomes are $Y_i(0) = \varepsilon_i$ and $Y_i(1) = f(v_i) + \varepsilon_i$ so $Y_i(1) - Y_i(0) = f(v_i)$ is the treatment effect. The individual-specific residuals $v_i$ and $\varepsilon_i$ are allowed to be arbitrarily correlated. Let $\beta_k$ denote the local average treatment effect (LATE) when comparing judge $k$ to the base judge: for instance, when $\lambda_k > 1/2$, $\beta_k = \frac{1}{\lambda_k - 1/2}\int_{1/2}^{\lambda_k} f(v)dv$. The $\beta_k$ values for the 4 groups of judges are also given in Table 1. The function $f(v)$ that delivers these parameters and further details of this example are in Appendix A.2.

The $\lambda_k$ and $\beta_k$ values are parameterized by objects $s$ and $h$, which control the IV strength and heterogeneity in the model respectively. The impact of these parameters are better illustrated in Figure 1 that plots the point masses for the four groups of judges in reduced-form. Parameter $s$ controls how far $E[X \mid Z]$ are spread across judges, which then affects the instrument strength. Parameter $h$ controls the distance between the mass points and a line with slope $\beta$ — this slope is the object of interest. If the impact of $X$ on $Y$ is homogeneous, then $h = 0$, and all mass points

Table 1: Parameter Values for Simple Example

| | | | | | |
|---|---|---|---|---|---|
| $\lambda_k$ | $\frac{1}{2} - s$ | $\frac{1}{2} - \frac{1}{2}s$ | $\frac{1}{2}$ | $\frac{1}{2} + \frac{1}{2}s$ | $\frac{1}{2} + s$ |
| $\beta_k$ | $\beta - \frac{h}{s}$ | $\beta + 2\frac{h}{s}$ | NA | $\beta - 2\frac{h}{s}$ | $\beta + \frac{h}{s}$ |

Figure 1: IV Strength and Heterogeneity in Reduced Form




*must* lie on a line — this implication is falsifiable by the data.

The simulation designs vary the values of $s$ and $h$ through the following parameters:

$$C_S = \sqrt{K}(c-1)s^2, \text{ and } C_H = \sqrt{K}(c-1)h^2. \tag{2}$$

Using Staiger and Stock (1997) asymptotics, $C_S$ is the parameter that determines whether there is strong or weak identification. Where $C$ is some positive arbitrary constant, $C_S \to \infty$ is an environment with strong identification where the object of interest can be estimated consistently, and $C_S \to C < \infty$ is an environment with weak identification where no consistent estimator exists.

For every design, I generate data under the null 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 either weak instruments or heterogeneity, and procedures that use a constructed instrument. The results are presented in Table 2, which I will refer to in the remainder of this section.

## 2.2 Issue with Weak Instruments

If we simply run the TSLS $t$-test for an over-identified model, then the estimator can be asymptotically biased and inference is invalid, a fact already known in the literature. This fact is also evident in Table 2, where TSLS has 100% rejection in many designs. 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 $Y$

Table 2: Rejection rates under the null for nominal size 0.05 test

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 0.900 | 0.066 | 1.000 | 0.078 | 0.100 | 0.101 | 0.055 | 0.053 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 1.000 | 0.033 | 1.000 | 0.248 | 0.076 | 0.271 | 0.045 | 0.045 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 0.998 | 0.024 | 1.000 | 0.264 | 0.055 | 0.297 | 0.051 | 0.048 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 0.996 | 0.066 | 1.000 | 0.036 | 0.043 | 0.044 | 0.039 | 0.048 |
| $C_H = 3, C_S = 3$ | 1.000 | 0.101 | 1.000 | 0.084 | 0.181 | 0.141 | 0.056 | 0.057 |
| $C_H = 3, C_S = 0$ | 1.000 | 0.141 | 1.000 | 0.107 | 0.242 | 0.192 | 0.069 | 0.054 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 1.000 | 0.145 | 0.048 | 0.058 | 0.074 | 0.074 | 0.064 | 0.052 |
| $C_H = 0, C_S = 2$ | 1.000 | 0.248 | 0.043 | 0.048 | 0.217 | 0.105 | 0.046 | 0.057 |
| $C_H = 0, C_S = 0$ | 1.000 | 0.378 | 0.044 | 0.058 | 0.337 | 0.128 | 0.064 | 0.050 |

Notes: The table displays rejection rates of various procedures (in columns) for various designs (in rows). Details of the data generating process are in Appendix A.2. Simulations use $K = 400, c = 5, \beta = 0$ with 1000 simulations. TSLS implements the standard two-stage-least-squares $t$-test for an over-identified IV system. EK implements the procedure in Evdokimov and Kolesár (2018). $T_{ee}$(MS) uses the statistic in Mikusheva and Sun (2022) with an oracle variance of their statistic. $T_{eX}$(MO) uses the $T_{eX}$ statistic with the variance estimator proposed in Matsushita and Otsu (2022). $\tilde{X}$-t uses a constructed instrument and runs TSLS for a just-identified IV system. $\tilde{X}$-AR uses the Anderson and Rubin (1949) (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.

on $\hat{X}$. With constant treatment effects, the asymptotic bias of TSLS depends on $\sum_i \varepsilon_i \hat{X}_i / \sum_i \hat{X}_i^2$. When every judge only has $c = 5$ cases, the influence of $v_i$ on $\hat{\pi}_{k(i)}$ and hence $\hat{X}_i$ is non-negligible. Since $\varepsilon_i$ and $v_i$ can be arbitrarily correlated, the numerator is biased. If the instruments are weak such that the denominator $\sum_i \hat{X}_i^2$ does not diverge sufficiently quickly, then the asymptotic bias can be large. Hence, the estimation problem arises from using $X_i$ to estimate $\hat{\pi}$.

A natural solution to address the bias in the TSLS estimator is to use the JIVE to estimate $\beta$. 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 $\hat{\pi}_{-i}$ the leave-one-out (L1O) coefficient. With $P = Z\left(Z'Z\right)^{-1} Z'$ 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:

$$\hat{\beta} = \frac{\sum_i Y_i \left(\sum_{j \neq i} P_{ij} X_j\right)}{\sum_i X_i \left(\sum_{j \neq i} P_{ij} X_j\right)}. \tag{3}$$

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 in the just-identified environment (Mikusheva and Sun, 2022; Yap, 2023) — it is a ratio of two normally distributed random variables. It is well-known that, in the just-identified IV context with weak IV, the rejection rate of the standard $t$-statistic can be up to 100% for a nominal 5% test (e.g., Dufour (1997); Staiger and Stock (1997)). Hence, like the just-identified IV context, by using a structural model that has sufficiently weak instruments and high covariance, the simulation can deliver high rejection rates.

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 2, where, with a sufficiently large correlation in the individual unobservables, rejection rates can be large. Further, Example 1 in Appendix A.3 can yield 100% rejection under the null (see Table 8). 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 $c = 5$.

**Remark 1.** In the literature, there have been several definitions of weak instruments in this context, which I clarify in this remark. Using Equation (2), there are three asymptotic regimes, ordered from the strongest to the weakest: (i) $\frac{1}{\sqrt{K}} C_S \to \infty$, (ii) $C_S \to \infty$, and (iii) $C_S \to C < \infty$. Regime (i) is a necessary condition for the TSLS estimator to be consistent, so $\frac{1}{\sqrt{K}} C_S \to C < \infty$ is what Stock and Yogo (2005) would refer to as weak instruments. Regime (ii) is a necessary condition for the JIVE to be consistent (e.g., Chao et al. (2012); Evdokimov and Kolesár (2018)). Regime (iii) is where no estimator is consistent (e.g., Mikusheva and Sun (2022)). If $K$ is fixed, then (i) and (ii) are the same asymptotically, and (iii) is the relevant weak identification asymptotic regime. If $K \to \infty$, then there is more ambiguity in what weakness means: Chao et al. (2012) and Evdokimov and Kolesár (2018) who assume (ii) are robust to weak instruments when defined in the Stock and Yogo (2005) sense, because $s$ can converge to 0, albeit at a slower rate than $\sqrt{K}$. In this paper, I follow the Staiger and Stock (1997) standard of weak identification where no consistent estimator exists, which corresponds to (iii) that EK18 is not robust to.

### 2.3 Issue with Heterogeneity

Next, consider proposals for inference that are developed for contexts with many weak IV. 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 $\beta_0$ 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 $E[Z_i e_i] = 0$.[1] Then, $T_{ee}$ is the L1O analog for the quadratic form that tests the moment $E[Z_i e_i] = 0$. Since observations are independent, the critical value for the test is obtained from a mean-zero normal distribution. In this model, $E[T_{ee}] = \sqrt{K}(c-1)h^2$.[2] Hence, when there are constant treatment effects such that

[1] An equivalent way to see how heterogeneity affects inference is through the framework of Hall and Inoue (2003) and Lee (2018): $E[Z_i e_i] = 0$ is a special case of a misspecified over-identified GMM problem. The instruments are individually valid, but every component of the $K$ moments in $E[Z_i e_i] = 0$ identifies a different treatment effect, so there is no parameter that satisfies all moments simultaneously under heterogeneity. Then, while the estimand is still interpretable as a combination of these treatment effects due to how GMM weights these moments, there are additional components in the variance that affect inference.

[2] This result can be obtained as a special case of Theorem 1 in Section 3 and using the fact that $\sum_i \sum_{j \neq i} P_{ij}^2 = \sum_i \sum_{j \neq i} (1/c^2) = \sum_i \frac{c-1}{c^2} = \sum_k \frac{c-1}{c}$.

$h = 0$ 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 $\beta_0$ such that $E[T_{ee}] = 0$, as shown in Lemma 3 of Appendix A.2. 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) (MO22) who use the statistic $T_{eX} = \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} e_i X_j = \frac{1}{\sqrt{K}} \sum_i e_i \tilde{X}_i$. This statistic can be interpreted as the LM (or score) statistic that uses the moment $E[e\tilde{X}] = 0$. They propose the following variance estimator $\hat{\Psi}_{MO}$:

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

While $T_{eX}$ has zero mean under heterogeneity, a result shown later in Section 3, the MO22 variance estimator was constructed under constant treatment effects, so the variance estimand differs from the true variance. It is shown in Appendix A.1 that $E\left[\hat{\Psi}_{MO}\right] \neq \mathrm{Var}\left(T_{eX}\right)$, and $\hat{\Psi}_{MO}$ 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 2 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-IV environment. With a single IV, the Anderson and Rubin (1949) (AR) 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 $t$-statistic inference is used as if $\tilde{X}_i$ is the instrument, then its rejection rates in designs with heterogeneity are usually higher than rejection rates of EK18 that accounts for the variance accurately, by comparing the $\tilde{X}$-t and EK columns in Table 2. Consequently, in the cases where EK under-rejects, $\tilde{X}$-t can have close to nominal rejection rates by coincidence.

Even if the weak IV robust AR procedure for just-identified IV were used, there can still be distortion in inference (see the $\tilde{X}$-AR column of Table 2). To see how the distortion arises, the AR $t$-statistic is $t_{\tilde{X}AR} := \sum_i e_i \tilde{X}_i / \sqrt{\hat{V}}$, where $\hat{V} = \sum_i \tilde{X}_i^2 \hat{\varepsilon}_i^2 / \left( \sum_i \tilde{X}_i^2 \right)^2$ and $\hat{\varepsilon}_i = e_i -$

$\tilde{X}_i \left(\sum_i e_i \tilde{X}_i\right) / \left(\sum_i \tilde{X}_i^2\right)$. Even though $t_{\tilde{X}AR}$ 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 estimand is $E\left[\sum_i \tilde{X}_i^2 e_i^2\right]$, whose expression is derived in Appendix A.2, and it does not converge to the true variance derived in Section 3 in general. Hence, using the just-identified AR procedure with a constructed instrument results in over-rejection. There are several papers that cluster standard errors by judges, but this approach faces a similar issue — details of this discussion are relegated to Appendix A.2.

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 $i = 1, \ldots, n$, we observe each individual's scalar outcome $Y_i$, scalar endogenous variable $X_i$, instrument $Z_i$, and covariates $W_i$, with $\dim(Z_i) = K$. 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 $R_i := E[X_i \mid Z_i, W_i]$ and $R_{Yi} := E[Y_i \mid Z_i, W_i]$ be linear in $Z_i$ and $W_i$. The model, written in the reduced-form and first-stage equations, is:

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

The setup implicitly conditions on $Z_i, W_i$, so $R_i, R_{Yi}$ are nonrandom.[3] Linearity in $Z$ and $W$ is not necessarily restrictive when there is full saturation or when $K$ is large.[4]

Define $e_i := Y_i - X_i\beta$, where $\beta$ is some estimand of interest, and $e_i$ is a linear transformation. Let $R_{\Delta i} := R_{Yi} - R_i\beta$ and $\nu_i := \zeta_i - \eta_i\beta$. These definitions imply $e_i = R_{\Delta i} + \nu_i$ 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 (i.e., projection) matrix

[3] If 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.

[4] 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.

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 $M = I - P$. $C$ denotes 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 $R_{\Delta i} = Z_i'(\pi_Y - \pi\beta) + W_i'(\gamma_Y - \gamma\beta)$ 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 $R_\Delta$ 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:

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

where $G$ is an $n \times n$ matrix that can take several forms. As the leading cases, if there are no covariates, using the projection matrix $G = H_Z = P$ is the standard JIVE, and when there are covariates, I use the unbiased JIVE "UJIVE" (Kolesár, 2013) with $G = (I - \text{diag}(H_Q))^{-1} H_Q - (I - \text{diag}(H_W))^{-1} H_W$.[5] In an environment with a binary instrument and many covariates interacted with the instrument, the saturated estimand "SIVE" (Chao et al., 2023; 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 when using JIVE without covariates or UJIVE with covariates with a saturated regression.[6] (Evdokimov and Kolesár, 2018) 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.

This paper restricts its attention to the following statistics:

$$(T_{ee}, T_{eX}, T_{XX})' := \frac{1}{\sqrt{K}} \sum_i \sum_{j\neq i} G_{ij} (e_i e_j, e_i X_j, X_i X_j)'. \tag{5}$$

These $T$ objects are observed because the $e_i$ objects can be calculated by using the null-imposed $\beta$. 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.

[5]Even 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.

[6]In the judge example without covariates above, we have $G = P$ and $\pi_{Yk} = \beta_k \pi_k$ where $\beta_k$ 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.

$T_{XX}$ corresponds to a first-stage F statistic that can be used as a diagnostic for weak instruments.

The asymptotic behavior depends on the following object:

$$r_n := \sum_i \left( \sum_{j \neq i} G_{ij} R_j \right)^2 + \sum_i \left( \sum_{j \neq i} G_{ij} R_{\Delta j} \right)^2 + \sum_i \sum_{j \neq i} G_{ij}^2. \tag{6}$$

Asymptotic theory in this paper uses $r_n/\sqrt{K} \to \infty$, which nests the environments of EK18, MS22, and MO22. EK18 assume $\sum_i \left( \sum_{j \neq i} G_{ij} R_j \right)^2 / \sqrt{K} \to \infty$, which implies $r_n/\sqrt{K} \to \infty$. The condition that $\sum_i \left( \sum_{j \neq i} G_{ij} R_j \right)^2 / \sqrt{K} \to \infty$ implies strong identification, but $r_n/\sqrt{K} \to \infty$ can also be achieved if either of the latter terms in Equation (6) diverges. MS22 and MO22 assume $K \to \infty$. Without covariates, $G = P$, so $\sum_i \sum_{j \neq i} G_{ij}^2 = O(K)$, and hence $r_n/\sqrt{K} \to \infty$. Hence, to apply the asymptotic theory in this paper, it suffices to have either strong identification, or $K \to \infty$. The only case ruled out is where $K$ is fixed, *and* there is weak identifcation in that $\sum_i \left( \sum_{j \neq i} G_{ij} R_j \right)^2 / \sqrt{K}$ does not diverge.

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

(c) There exists $\underline{c} > 0$ such that for any $c_1, c_2, c_3$ that are not all 0,
$\frac{1}{r_n} \sum_i \left( c_3 \sum_{j \neq i} (G_{ij} + G_{ji}) R_j + c_2 \sum_{j \neq i} G_{ji} R_{\Delta j} \right)^2 + \frac{1}{r_n} \sum_i \left( c_1 \sum_{j \neq i} (G_{ij} + G_{ji}) R_{\Delta j} + c_2 \sum_{j \neq i} G_{ij} R_j \right)^2 + \frac{1}{r_n} \sum_i \sum_{j \neq i} \left( G_{ij}^2 + G_{ij} G_{ji} \right) \geq \underline{c}$.

(d) $\frac{1}{r_n^2} \sum_i \left( \left( \sum_{j \neq i} G_{ij} R_j \right)^4 + \left( \sum_{j \neq i} G_{ij} R_{\Delta j} \right)^4 + \left( \sum_{j \neq i} G_{ji} R_j \right)^4 + \left( \sum_{j \neq i} G_{ji} R_{\Delta j} \right)^4 \right) \to 0$.

(e) $||\frac{1}{r_n} G_L G_L'||_F + ||\frac{1}{r_n} G_U G_U'||_F \to 0$, 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. 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 $||G_L G_L'||_F \leq C\sqrt{K}$. Mechanically, if there is weak IV and fixed K, then $||\frac{1}{r_n} G_L G_L'||_F = \frac{1}{K} O(\sqrt{K}) \neq o(1)$, so (e) fails when $r_n/\sqrt{K}$

does not diverge. Notably, the conditions do not require $P_{ii} \to 0$ so the $\pi, \pi_Y$ 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,*

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

*and for some variance matrix $V$, as $r_n / \sqrt{K} \to \infty$,*

$$V^{-1/2} \begin{pmatrix} T_{ee} - \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_{\Delta i} R_{\Delta j} \\ T_{eX} \\ T_{XX} - \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} R_i R_j \end{pmatrix} \xrightarrow{d} N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, I_3 \right). \tag{7}$$

Theorem 1 states a numerical equivalence on the difference between $\hat{\beta}_{JIVE}$ and $\beta_{JIVE}$. $S_K$ is the concentration parameter corresponding to the instrument strength. In the model of Section 2, the mapping to the reduced-form $\pi$ can be found in Appendix A.2, so the concentration parameter is given by $S_K = \frac{1}{\sqrt{K}} \sum_k (c-1) \pi_k^2 = \frac{5}{8} \sqrt{K} (c-1) s^2$. If the instruments are strong, then $S_K \to \infty$, 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 the leave-one-out sums as a martingale difference array, and then applying the martingale CLT. 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 for $G = P$, but $G$ for UJIVE is not symmetric in general. 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 $t$-statistic. While Yap (2023) and MS22 have observed this result with many weak IV, 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 $T_{eX} / \sqrt{\mathrm{Var}(T_{eX})}$ 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 $\beta_0$ and an estimator for the variance of $\sqrt{K} T_{eX}$, $\hat{V}_{LM}$, defined later in this section. Then, reject if $K T_{eX}^2 / \hat{V}_{LM} \geq \Phi \left(1 - \alpha/2\right)^2$ for a size $\alpha$ test where $\Phi(.)$ 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 Appendix B that, for $V_{LM} := \text{Var}\left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j\right)$,

$$
\begin{aligned}
V_{LM} = &\sum_i \sum_{j \neq i} \sum_{k \neq i} E\left[\nu_i^2\right] G_{ij} G_{ik} R_j R_k + \sum_i \sum_{j \neq i} G_{ij}^2 E\left[\nu_i^2\right] E\left[\eta_j^2\right] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E\left[\eta_i \nu_i\right] E\left[\eta_j \nu_j\right] \\
&+ 2 \sum_i \sum_{j \neq i} \sum_{k \neq i} E\left[\nu_i \eta_i\right] G_{ij} G_{ki} R_j R_{\Delta k} + \sum_i \sum_{j \neq i} \sum_{k \neq i} E\left[\eta_i^2\right] G_{ji} G_{ki} R_{\Delta j} R_{\Delta k}.
\end{aligned} \tag{8}
$$

With constant treatment effects, only the first line appears in the variance as $R_\Delta = 0$. With $G = P$, 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 Sølvsten (2023) (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. The variance estimator is:

$$
\hat{V}_{LM} := A_1 + A_2 + A_3 + A_4 + A_5, \tag{9}
$$

with

$$
\begin{aligned}
A_1 &:= \sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ik} X_k e_i \left(e_i - Q_i' \hat{\tau}_{\Delta,-ijk}\right), \\
A_2 &:= 2 \sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ij} X_j G_{ki} e_k e_i \left(X_i - Q_i' \hat{\tau}_{-ijk}\right), \\
A_3 &:= \sum_i \sum_{j \neq i} \sum_{k \neq i} G_{ji} e_j G_{ki} e_k X_i \left(X_i - Q_i' \hat{\tau}_{-ijk}\right), \\
A_4 &:= -\sum_i \sum_{j \neq i} \sum_{k \neq j} G_{ji}^2 X_i \check{M}_{ik,-ij} X_k e_j \left(e_j - Q_j' \hat{\tau}_{\Delta,-ijk}\right), \\
A_5 &:= -\sum_i \sum_{j \neq i} \sum_{k \neq j} G_{ij} G_{ji} e_i \check{M}_{ik,-ij} X_k e_j \left(X_j - Q_j' \hat{\tau}_{-ijk}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\tau}_{-ijk} &:= \left(\sum_{l \neq i,j,k} Q_l Q_l'\right)^{-1} \sum_{l \neq i,j,k} Q_l X_l, \\
\hat{\tau}_{\Delta,-ijk} &:= \left(\sum_{l \neq i,j,k} Q_l Q_l'\right)^{-1} \sum_{l \neq i,j,k} Q_l e_l,
\end{aligned}
$$

$$D_{ij} := M_{ii}M_{jj} - M_{ij}^2, \text{ and}$$

$$\check{M}_{ik,-ij} := \frac{M_{jj}M_{ik} - M_{ij}M_{jk}}{D_{ij}} = -Q_i' \left( \sum_{l \neq i,j} Q_l Q_l' \right)^{-1} Q_k.$$

Following AS23, I make an assumption to ensure that the L3O estimator is well-defined.[7]

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

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

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 Appendix A.1. 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 Appendix A.1,* $E\left[\hat{V}_{LM}\right] = V_{LM}$ *and the variance estimator is consistent, i.e.,* $\hat{V}_{LM}/V_{LM} \xrightarrow{p} 1$.

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, but this approach no longer works when these estimated coefficients are inconsistent. To be precise, applying Chebyshev's inequality for any $\epsilon > 0$ yields:

$$\Pr\left( \left| \frac{\hat{V}_{LM} - V_{LM}}{V_{LM}} \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \frac{Var\left(\hat{V}_{LM}\right)}{V_{LM}^2} + \frac{1}{\epsilon^2} \frac{\left( E\left[\hat{V}_{LM}\right] - V_{LM} \right)^2}{V_{LM}^2}. \tag{10}$$

Without an unbiased estimator and when reduced-form coefficients cannot be consistently estimated, the second term in (10) 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 (10) 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. The leave-three-out (L3O) approach has this unbiasedness property for linear regressions: when leaving three observations out in the inner-most sum of the $A$ expressions, the estimated coefficient $\hat{\tau}_{-ijk}$ is independent of $i, j, k$ and is unbiased for $\tau$. Then, when taking the expectation through a product of random variables of $i, j, k$ and $\hat{\tau}_{-ijk}$, $\tau$ can be used in place of the $\hat{\tau}_{-ijk}$ component, and the

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

expectations of individual components can be isolated. For instance,

$$E\left[\sum_i \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}X_jG_{ik}X_ke_i\left(e_i - Q_i'\hat{\tau}_{\Delta,-ijk}\right)\right] = \sum_i \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}E\left[X_j\right]G_{ik}E\left[X_k\right]E\left[e_i\left(e_i - Q_i'\hat{\tau}_{\Delta,-ijk}\right)\right]$$
$$= \sum_i \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}R_jG_{ik}R_kE\left[\nu_i^2\right], \tag{11}$$

which recovers the triple sums in the $V_{LM}$ expression of Equation (8). Without leaving out observations $j$ and $k$, we would not be able to isolate $E[X_j]$ and $E[X_k]$ in the first equality. Similarly, without leaving out observation $i$, we would not be able to isolate $\tau_\Delta$ on expectation to obtain $E[\nu_i^2]$ in the second equality. An analogous argument applies to other components of $V$ in (7). Assuming that the residuals have zero mean conditional on $Q$ is crucial: if we merely have $E[Q\zeta] = 0$, this argument can no longer be applied.

**Remark 3.** While the proposed $\hat{V}_{LM}$ is motivated by AS23, the contexts and estimators are different. First, the statistic that we are estimating the variance for is different: AS23 demeaned their $\mathcal{F}$ statistic using $\hat{E}_\mathcal{F}$, where $\hat{E}_\mathcal{F}$ is estimated using L1O, so they are interested in the variance of $\mathcal{F} - \hat{E}_\mathcal{F}$ that is mean zero; I use a mean-zero L1O statistic directly in $T_{eX}$. Second, the expectation of their variance estimator takes the form of their (9), which is analogous to the sum of $A_1$ and $A_4$ using the notation above, so repeated applications of their estimator is insufficient to recover all five terms exactly. Hence, to adjust for the $A_4$ and $A_5$ terms here, I additionally require another estimator, and its form is similarly motivated by a L3O reasoning.

# 4 Power Properties

This section characterizes power properties of the valid LM procedure. 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 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.

## 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 $(\eta, \zeta)$ in the model of

Section 3.1 are jointly normal with known variance. To be precise,

$$\begin{pmatrix} \zeta_i \\ \eta_i \end{pmatrix} \sim N\left(0,\Omega\right) = N\left(0, \begin{bmatrix} \omega_{\zeta\zeta} & \omega_{\zeta\eta} \\ \omega_{\zeta\eta} & \omega_{\eta\eta} \end{bmatrix}\right). \tag{12}$$

Define:

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

I restrict attention to tests that are invariant to rotations of $Z$, i.e., transformations of the form $Z \to ZF'$ 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 $F$. 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 $Z \to ZF'$ 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*

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

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.

**Proposition 1.** *With Equation (12), $K \to \infty$ and $\frac{1}{\sqrt{K}}\left(\pi_Y'Z'Z\pi_Y, \pi'Z'Z\pi_Y, \pi'Z'Z\pi\right) \to (C_{YY}, C_Y, C_S)$,*

$$\frac{1}{\sqrt{K}} \begin{pmatrix} s_1's_1 - K\omega_{\zeta\zeta} - C_{YY} \\ s_1's_2 - K\omega_{\zeta\eta} - C_Y \\ s_2's_2 - K\omega_{\eta\eta} - C_S \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma\right) \tag{13}$$

*for some variance matrix $\Sigma$. 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 $K \to \infty$ because objects like $s_1's_1$ can be written as a sum of $K$ objects. With an appropriate representation to obtain independence, a CLT 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 (13). Without covariates such that $G = P$, I define $(T_{YY}, T_{YX}, T_{XX}) := \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij}(Y_iY_j, Y_iX_j, X_iX_j)$. This $(T_{YY}, T_{YX}, T_{XX})$ is the L1O analog of the maximal invariant $(s_1's_1, s_1's_2, s_2's_2)$, which relates to JIVE directly because $\hat{\beta}_{JIVE} = T_{YX}/T_{XX}$.[8] 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.[9] Since $(T_{YY}, T_{YX}, T_{XX})$ is the L1O analog and has the same distribution as the maximal invariant, I restrict our attention to tests that are functions of $(T_{YY}, T_{YX}, T_{XX})$.

While validity results in Section 3 apply even when $K$ is small, the optimality results here do not apply. Based on Proposition 1, the distribution of the maximal invariant is approximately normal when $K$ is large. When $K$ is fixed, the distribution of the maximal invariant is different from the distribution of L1O statistics, and focusing on the L1O statistics is not justified.

### 4.2 Discussion of Asymptotic Problem

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

$$\begin{pmatrix} T_{YY} \\ T_{YX} \\ T_{XX} \end{pmatrix} \sim N(\mu, \Sigma), \mu = \begin{pmatrix} \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_{Yj} \\ \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_j \\ \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} P_{ij} R_i R_j \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \cdot & \sigma_{22} & \sigma_{23} \\ \cdot & \cdot & \sigma_{33} \end{pmatrix}. \tag{14}$$

There are several restrictions in the $\mu$ vector, which is assumed to be finite. Since $P$ is a projection matrix, $\sum_i \sum_{j \neq i} P_{ij} R_i R_j = \sum_i R_i (\sum_j P_{ij} R_j - P_{ii} R_i) = \sum_i M_{ii} R_i^2$. Since the annihilator matrix $M$ has positive entries on its diagonal, we obtain $\mu_3 \geq 0$ and a similar argument yields $\mu_1 \geq 0$. With $\mu_2 = \sum_i \sum_{j \neq i} P_{ij} R_{Yi} R_j = \sum_i M_{ii} R_{Yi} R_i$, the Cauchy-Schwarz inequality implies $\mu_2^2 \leq \mu_1 \mu_3$. Constant treatment effects implies $\mu_2^2 = \mu_1 \mu_3$, which is a special case of the environment here. Even with covariates, if the regression is fully saturated with $G$ given by UJIVE, Proposition 4 in Appendix A.1 shows that the same inequality restrictions hold. 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 $\Sigma$ places further restrictions on $\mu$, which can give more information about $\beta_{JIVE} = \mu_2/\mu_3$. While $\Sigma$ 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 $\beta$ features in $\Sigma$. With more structure, there can be more restrictions on $\mu$, but if there is no structural model where $\Sigma$ is uninformative, then any necessary restriction should be accounted for in the asymptotic problem. Hence, Appendix A.3.1 establishes that there exists a structural model where $\Sigma$ is uninformative about $\mu$, and $\mu_1, \mu_3 \geq 0$, so $\mu_2^2 \leq \mu_1 \mu_3$ are the *only* restrictions on $\mu$.[10] While the result establishes that there exists a structural model where there are no further restrictions, for any given structural model, there can still be further restrictions.

[8] To see this analogy, $s_1' s_2 = Y'Z(Z'Z)^{-1}Z'X = Y'PX = \sum_i \sum_j P_{ij} Y_i X_j$.

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

[10] Since the model in Section 2 is binary, it is insufficient for such a general result, and a continuous $X$ is required, so the example is relegated to Appendix A.3.1.

### 4.3 Analytic Results

Using the asymptotic problem of Equation (14), testing $H_0 : \mu_2/\mu_3 = \beta^*$ is identical to testing $H_0 : \mu_2 - \beta^*\mu_3 = 0$. Since $\beta^*$ 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 $\mu_2 = 0$, and $(T_{YY}, T_{YX}, T_{XX}) = (T_{ee}, T_{eX}, T_{XX})$. Let $\mu^A$ denote the mean under the alternative and $\mu^H$ under the null. The remainder of this section presents theoretical results for power, and numerical results beyond the environment covered by theory are relegated to Appendix A.3.2.

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

The one-sided test is the most powerful test for testing against a particular subset of alternatives $\mathcal{S} := \left\{ \left(\mu_1^A, \mu_2^A, \mu_3^A\right) : \mu_1^A - \frac{\sigma_{12}}{\sigma_{22}}\mu_2^A \geq 0, \mu_3^A - \frac{\sigma_{23}}{\sigma_{22}}\mu_2^A \geq 0 \right\}$. 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. The proposition makes no statement about alternative hypotheses that are not in $\mathcal{S}$. A more powerful test can be constructed when $\mu_2^A$ is large and covariance $\sigma_{23}, \sigma_{12}$ are large.

**Proposition 2.** *The one-sided LM test is the most powerful test for testing any alternative hypothesis* $\left(\mu_1^A, \mu_2^A, \mu_3^A\right) \in \mathcal{S}$ *in the asymptotic problem of Equation* (14).

For a given $\left(\mu_1^A, \mu_2^A, \mu_3^A\right)$ 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 $\mu^H$ 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 $\mu^H$ is in the null space, so LM is the most powerful test. 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.

Turning to two-sided tests, I consider the theoretical benchmark of a uniformly most powerful unbiased test (e.g., Lehmann and Romano (2005); Moreira (2009b)).

**Proposition 3.** *Consider a restriction of the alternative $\mu$ 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* (14).

The argument for optimality applies a standard optimality result from Lehmann and Romano (2005) on the exponential family, which includes the normal distribution. To apply the Lehmann and Romano (2005) result, we require a convex parameter space and the the existence of alternative values above and below the null value.[11] It can be verified that the restricted parameter space is still

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

convex, and the restriction to the interior ensures the latter condition is satisfied. 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).

**Remark 4.** 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 $\mu_1$ and $\mu_3$, then a 97% test for LM. (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 the algorithm from Elliott et al. (2015) can also be applied by using a grid on structural parameters.

Studying optimality in the over-identified IV environment has thus far been complicated. With constant treatment effects, both $s_1' s_1$ and $s_1' s_2$ are informative of the object of interest $\beta$, because constant treatment effects implies $\mu_1 = \beta^2 \mu_3$. However, once we impose $\mu_1 > 0$ under the null that $\beta = 0$, we rule out constant treatment effects by focusing on the interior of the alternative space. Then, the statistic associated with $\mu_1$ is no longer directly informative of $\beta$. Imposing heterogeneity is hence the key to obtaining this UMPU result.

# 5 Implementation

Expressions for the test are given in Section 3, which can be efficiently implemented using matrix operations. Inverting the test to obtain a confidence set is also straightforward in this procedure, as the bounds of the confidence set are derived in closed-form in this section.

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 (9) so that they are written in terms of $X$ and $Y$. The two-sided test rejects: $\left(\sum_i \sum_{j \neq i} G_{ij} e_i X_j\right)^2 / \hat{V}_{LM} \geq q = \Phi(1-\alpha/2)^2$. Let $T_{YX} := \frac{1}{\sqrt{K}} \sum_i \sum_{j \neq i} G_{ij} Y_i X_j$. Then, $\sum_i \sum_{j \neq i} G_{ij} e_i X_j = \sqrt{K}\,(T_{YX} - T_{XX}\beta_0)$, so squaring it results in a term that is quadratic in $\beta_0^2$. With $\hat{V}_{LM} = B_0 + C_1 \beta_0 + B_2 \beta_0^2$ quadratic in $\beta_0$, where $B_0, B_1, B_2$ are derived in Appendix D, 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)). Calculations for coefficients is similar to that of the L3O variance.

**Lemma 2.** *The test does not reject when* $\left(KT_{XX}^2 - qB_2\right)\beta_0^2 - \left(2KT_{YX}T_{XX} + qB_1\right)\beta_0 + \left(KT_{YX}^2 - qB_0\right) \leq 0$. *Let* $D := \left(2KT_{YX}T_{XX} + qB_1\right)^2 - 4\left(KT_{XX}^2 - qB_2\right)\left(KT_{YX}^2 - qB_0\right)$. *If* $D \geq 0$ *and* $KT_{XX}^2 - qB_2 \geq 0$, *then the upper and lower bounds of confidence set are:*

$$\frac{\left(2KT_{YX}T_{XX} + qB_1\right) \pm \sqrt{D}}{2\left(KT_{XX}^2 - qB_2\right)}.$$

*If* $D < 0$ *and* $KT_{XX}^2 - qB_2 < 0$, *then the confidence set is empty. Otherwise, the confidence set is unbounded.*

Due to $+qB_1, -qB_2$ in the expression of the upper and lower bounds, the confidence set is not necessarily centered around $\hat{\beta}_{JIVE} = T_{YX}/T_{XX}$. In the numerical illustrations that follow, I fully saturate the regressions with covariates as doing so simplifies implementation: when $G$ is block-diagonal, it suffices to loop over blocks, which significantly speeds up the code.

# 6 Numerical Illustrations

## 6.1 Simulations

The general model in Section 3 can be justified by several structural models. In this section, I focus on the simple example from Section 2. There are two sets of simulations that assess the size: I generate data under the null and assess how close the rejection rates of various procedures are to the nominal rate. One set of size simulations uses a large $K$ while the other a small $K$. I also report one set of simulations that assess power: I generate data under some alternative and assess the rejection rates across procedures. There are more simulation results using several different structural models in Appendix A.4, including settings with continuous treatment $X$, and with covariates. The results are qualitatively similar in those simulations, suggesting that the numerical findings are not unique to the data-generating process chosen.

Table 2 in Section 2 reports rejection rates under the null for a relatively large number of judges with $K = 400$, each with a small number of cases at $c = 5$. L3O performs well across various designs, while existing procedures can substantially over-reject in at least one design. The LMorc column is included as an infeasible theoretical benchmark that uses an oracle variance: this should have nominal size when normality holds because the variance is not estimated. The difference between LMorc and L3O is attributed to the variance estimation procedure.

Table 3 reports rejection rates under the null for a small number of judges with $K = 4$ and a large number of cases at $c = 200$. Based on the theory in Section 3, L3O should be valid when the instrument is strong, i.e., in the cases with $C_S = .5c$, which is what we observe. Notably, even when $C_S = 2$ or $C_S = 0$, the over-rejection for L3O is not too severe. EK performs very well in the cases with $C_S = .5c$ as expected in their theory. In contrast, MS and MO can over-reject severely with strong heterogeneity, even when instruments are strong.

Table 4 reports rejection rates under the alternative. When $C_S = 0$, the instrument should be completely uninformative about the true parameter, so we should have 0.05 rejection rate for a valid test, which is what we observe for L3O. When $C_S = 2\sqrt{K}$, all procedures, including L3O, are very informative. Considering the designs with $C_H = 0$ is most interesting, because this is an environment where MS and MO are valid, and the theoretical optimality result excludes this case. Looking at the case with $C_H = 0, C_S = 2$, L3O is less powerful than MO in small samples, but we should expect both procedures to converge to the same variance in larger samples. L3O is a lot less powerful than MS for $C_H = 0, C_S = 2$, suggesting that this data-generating process favors MS with constant treatment effects. L3O nonetheless outperforms EK and $\tilde{X}$-t in terms of power.

Table 3: Rejection rates under the null for nominal size 0.05 test

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = .5c, C_S = .5c$ | 0.325 | 0.050 | 1.000 | 0.316 | 0.324 | 0.323 | 0.056 | 0.051 |
| $C_H = .5c, C_S = 2$ | 0.575 | 0.012 | 1.000 | 0.820 | 0.277 | 0.833 | 0.040 | 0.047 |
| $C_H = .5c, C_S = 0$ | 0.501 | 0.005 | 1.000 | 0.856 | 0.335 | 0.881 | 0.061 | 0.050 |
| $C_H = 2, C_S = .5c$ | 0.082 | 0.057 | 0.604 | 0.081 | 0.073 | 0.082 | 0.065 | 0.060 |
| $C_H = 2, C_S = 2$ | 0.485 | 0.013 | 0.625 | 0.347 | 0.326 | 0.466 | 0.109 | 0.046 |
| $C_H = 2, C_S = 0$ | 0.461 | 0.011 | 0.624 | 0.341 | 0.349 | 0.497 | 0.107 | 0.047 |
| $C_H = 0, C_S = .5c$ | 0.064 | 0.045 | 0.043 | 0.044 | 0.046 | 0.051 | 0.055 | 0.043 |
| $C_H = 0, C_S = 2$ | 0.437 | 0.102 | 0.048 | 0.040 | 0.296 | 0.134 | 0.066 | 0.042 |
| $C_H = 0, C_S = 0$ | 0.590 | 0.181 | 0.049 | 0.029 | 0.431 | 0.163 | 0.059 | 0.045 |

Notes: $K = 4, c = 200$, and designs are otherwise identical to Table 2.

Table 4: Rejection rates under the alternative for nominal size 0.05 test

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 0.449 | 0.100 | 1.000 | 0.417 | 0.044 | 0.452 | 0.179 | 0.153 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 0.825 | 0.028 | 1.000 | 0.273 | 0.063 | 0.298 | 0.050 | 0.043 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $C_H = 3, C_S = 3$ | 0.322 | 0.491 | 1.000 | 0.865 | 0.150 | 0.889 | 0.737 | 0.752 |
| $C_H = 3, C_S = 0$ | 1.000 | 0.080 | 1.000 | 0.109 | 0.196 | 0.177 | 0.052 | 0.057 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $C_H = 0, C_S = 2$ | 0.881 | 0.400 | 0.978 | 0.747 | 0.075 | 0.812 | 0.692 | 0.752 |
| $C_H = 0, C_S = 0$ | 1.000 | 0.366 | 0.046 | 0.033 | 0.322 | 0.092 | 0.061 | 0.049 |

Notes: $K = 100, \beta = 0.1, c = 5$, and designs are otherwise identical to Table 2.

## 6.2 Returns to Education

Angrist and Krueger (1991) were interested in the impact of years of education (X) on log weekly wages (Y). They instrument for education using the quarter of birth (QOB). I implement UJIVE using full interaction of QOB with the state of birth and year of birth (resulting in 1530 instruments) without other controls, which is similar to Table VII(2) of Angrist and Krueger (1991) that uses the same set of controls but without full saturation. The implementation here differs from the implementation of MS and MO in that I do not linearly partial out other covariates, but merely saturate on state and year of birth.

To ensure that the different procedures are directly comparable, I adapt the MS and MO inference procedures to target UJIVE, so that the estimand is the same across all procedures and differing results can be attributed purely to inference. TSLS is consequently not a meaningful comparison as the estimand is different from the others. When using the MS procedure, I adapt their "naive" variance estimator instead of their debiased estimator, because the naive estimator is more easily adaptable to the UJIVE, and both estimators are consistent under the same asymptotic regime of weak identification.

The results are reported in Table 5. Being robust to weak and many IV results in L3O having a longer confidence interval than EK. With full saturation, MS and MO yield unbounded confidence sets, while L3O still yields a bounded confidence set, showing how robustness to heterogeneity changes the shape of the confidence set in this context. Using the notation from Section 5, observe that the shape of the confidence set depends on the coefficient on $\beta_0^2$. In particular, for $\Psi_2 := \frac{1}{K}\sum_i \left(\sum_{j\neq i} G_{ij}X_j\right)^2 X_i^2 + \frac{1}{K}\sum_{i\neq j}^n G_{ij}^2 X_i^2 X_j^2$, MO is unbounded when $T_{XX}^2 - q\Psi_2 < 0$ and L3O is unbounded when $T_{XX}^2 - qB_2 < 0$, where $q$ is 3.84 for a 5% test. Consequently, in this application, we can think of $T_{XX}^2/\Psi_2 = 0.102$ and $T_{XX}^2/B_2 = 11.8$ as first-stage statistics for MO and L3O respectively that determine whether the confidence sets are bounded.[12] Analogously, when solving a quartic equation in MS using their naive variance estimator, an unbounded set occurs here as $T_{XX}^2/\left(\frac{2}{K}\sum_{i\neq j}^n G_{ij}^2 X_i^2 X_j^2\right) = 0.0545 < q$, where the denominator is their coefficient on $\beta_0^4$ when inverting their test.

Due to the $\check{M}$ terms in the L3O expression, it is difficult to compare the estimates directly. However, it is possible to compare the estimands of these coefficients in the judge example without covariates. Due to Appendix A.1,

$$E\left[K\Psi_2\right] - E\left[B_2\right] = \sum_i M_{ii}R_i^2\left(R_i^2 - 3\left(1-2P_{ii}\right)E\left[\eta_i^2\right]\right). \tag{15}$$

If $R_i^2 > 3\left(1-2P_{ii}\right)E\left[\eta_i^2\right]$, then MO is more likely unbounded. Intuitively, the MO estimator contains additional products of $R$ that are not present in the true variance, and there are products of $R$ and the error present in the true variance that MO does not account for, motivating the

[12] These statistics are "F" statistics with different variance estimators. The MS and MO variance estimators converge to the same object under weak identification such that $\frac{1}{K}\sum_i\sum_{j\neq i} G_{ij}R_iR_j \to 0$, which is not imposed by the asymptotic regime in this paper.

Table 5: 95% Confidence Sets for Returns to Education

| | EK | $T_{ee}$(CMS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O |
|---|---|---|---|---|---|---|
| LB | 0.033 | -$\infty$ | -$\infty$ | 0.027 | 0.027 | 0.022 |
| UB | 0.173 | $\infty$ | $\infty$ | 0.179 | 0.189 | 0.210 |
| Estimate | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 | 0.103 |
| CIlength | 0.140 | $\infty$ | $\infty$ | 0.152 | 0.161 | 0.188 |

Notes: Estimate reports the UJIVE. The variance for $T_{ee}$ is calculated using their "naive" estimator earlier proposed by Crudu et al. (2021) (CMS). Procedures are otherwise identical to Table 2.

Table 6: 95% Confidence Sets for Misdemeanor Prosecution

| | EK | $T_{ee}$(CMS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O |
|---|---|---|---|---|---|---|
| LB | -0.151 | NA | -0.220 | -0.187 | -0.188 | -0.201 |
| UB | -0.076 | NA | -0.019 | -0.039 | -0.038 | -0.028 |
| Estimate | -0.113 | -0.113 | -0.113 | -0.113 | -0.113 | -0.113 |
| CIlength | 0.075 | NA | 0.201 | 0.148 | 0.150 | 0.173 |

Notes: Procedures are identical to Table 5.

aforementioned difference. We can interpret this condition as MO being more likely unbounded when the signal-to-noise ratio $R_i^2/E\left[\eta_i^2\right]$ is sufficiently large. In this application, by observing that the first-stage statistic of L3O is an order of magnitude larger than that of MO (i.e., $B_2$ is an order of magnitude smaller), and by comparing the MS and MO first-stage statistics, there is evidence that $R_i^2/E\left[\eta_i^2\right]$ is large.[13] This result does not depend on heterogeneity, because the coefficient of $\beta_0^2$ depends only on how $X$ is combined in the variance estimator.

### 6.3 Misdemeanor Prosecution

Agan et al. (2023) were interested in the effect of misdemeanor prosecution (X) on criminal complaint in two years (Y). They instrument for misdemeanor prosecution using the assistant district attorneys (ADAs) who decide if a case should be prosecuted in the Suffolk County District Attorney's Office in Massachusetts. As Agan et al. (2023) argued that as-if randomization holds conditional on court-by-time controls and that individual covariates are not required for relevance or exogeneity to hold in this context, the confidence set is constructed using full saturation of court-by-year and court-by-day-of-week fixed effects with no other controls for individual covariates.

[13] Comparing the statistics between MO and MS implies $\frac{1}{K}\sum_i\left(\sum_{j\neq i}G_{ij}X_j\right)^2X_i^2 < \frac{1}{K}\sum_i\sum_{j\neq i}G_{ij}^2X_i^2X_j^2$, which can equivalently be written as $\sum_i\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}X_jX_kX_i^2 < 0$. With full saturation, observations $i,j$ have $G_{ij} < 0$ when they are in the same covariate group but have different instrument values. Under the MS and MO asymptotic regimes where $\frac{1}{K}\sum_i\sum_{j\neq i}G_{ij}R_iR_j \to 0$ so $R_i^2/E[\eta_i^2]$ is negligible, we obtain $\sum_i\sum_{j\neq i}G_{ij}^2E[X_i^2]E[X_j^2] = \sum_i\sum_{j\neq i}G_{ij}^2\left(R_i^2+E[\eta_i^2]\right)\left(R_j^2+E[\eta_j^2]\right) = \sum_i\sum_{j\neq i}G_{ij}^2E[\eta_i^2]E[\eta_j^2]+o(1)$, and $\frac{1}{K}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}E\left[X_jX_kX_i^2\right] = \frac{1}{K}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}R_jR_k\left(R_i^2+E[\eta_i^2]\right) = o(1)$ is asymptotically negligible. Since the difference between MS and MO is the same magnitude as the MS statistic, $\frac{1}{K}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}E\left[X_jX_kX_i^2\right]$ is of similar order as $\sum_i\sum_{j\neq i}G_{ij}^2E[X_i^2]E[X_j^2]$, so $R_i^2/E[\eta_i^2]$ is non-negligible in this application.

As reported in Table 6, with full saturation, the UJIVE is $-0.11$, so not prosecuting decreases the probability of criminal involvement by 11 percentage points.[14] The L3O confidence interval (CI) is more than twice that of EK: unlike Section 6.2 where $n/K = 221$, we have $n/K = 11.9$ here, so a variance estimator that is robust to many IV has a larger impact on CI. MS has an empty confidence set while L3O has a bounded set, showing how being robust to heterogeneity can change conclusions. Mechanically, the confidence set for MS solves a quartic equation, so an empty set can occur, but it is difficult to characterize when this phenomenon occurs in general.

The L3O CI is also shorter than MO, so being robust to heterogeneity decreases the length of the CI. Considering how the length of the MO confidence set is longer than L3O while being oversized in simulations, there is a question of when MO is conservative. While it is difficult to compare the confidence intervals or variance estimators directly, it is possible to compare the null-imposed variance estimands in the judge example without covariates. Due to Appendix A.1,

$$E\left[\hat{\Psi}_{MO}\right] - Var\left(\sum_i \sum_{j\neq i} P_{ij} e_i X_j\right) = \sum_i M_{ii} R_{\Delta i}^2 \left(R_i^2 - (1-2P_{ii})E\left[\eta_i^2\right]\right) - 2\sum_i M_{ii}(1-2P_{ii})E\left[\eta_i \nu_i\right] R_i R_{\Delta i}.$$

Then, MO is conservative when: (i) $R_i^2 > (1-2P_{ii})E\left[\eta_i^2\right]$, and (ii) $E\left[\eta_i \nu_i\right]$ is negatively correlated with $R_i R_{\Delta i}$, when $P_{ii} < 1/2$. In (i), $R_{\Delta i}^2$ only affects the magnitude of the difference, and not the sign, so this condition can be interpreted as a condition on the signal-to-noise ratio as before. Condition (ii) results from the $\sum_i M_{ii}(1-2P_{ii})E\left[\eta_i \nu_i\right] R_i R_{\Delta i}$ term that MO does not account for, and covariances can be positive or negative in general.

# 7 Conclusion

This paper has documented how weak instruments and heterogeneity can interact to invalidate existing procedures in the environment of many instruments. Addressing both problems simultaneously, this paper contributes a feasible and robust method for valid inference. The procedure is shown to be valid as the limiting distribution of commonly used statistics, including the LM statistic, in an environment with many weak instruments and heterogeneity, is normal, and a leave-three-out variance estimator is consistent for obtaining the variance of the LM statistic. Further, the associated confidence set can be derived in closed form. Beyond its validity, the LM test is also optimal, as it is the uniformly most powerful unbiased test in the asymptotic distribution for the interior of the alternative space. In light of the broader econometric literature on the value of saturated regressions and how many instruments can arise from them, this paper presents a highly applicable, robust, and powerful inference procedure for IV.

[14]This result is smaller than $-0.36$ reported in their Table III(3) that uses TSLS with a leniency measure. The result is more similar to the UJIVE robustness check in their Table A.1(5) of $-0.15$ with full saturation of the instrument, but their specification includes case/ defendant covariates, which results in a different estimator.

## 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., V. Marmer, and Z. Yu (2019): "On optimal inference in the linear IV model," *Quantitative Economics*, 10, 457–485.

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. (2016): "Conditional linear combination tests for weakly identified models," *Econometrica*, 84, 2155–2182.

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*.

Bound, J., D. A. Jaeger, and R. M. Baker (1995): "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak," *Journal of the American statistical association*, 90, 443–450.

Cattaneo, M. D., M. Jansson, and W. K. Newey (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. Newey, and T. Woutersen (2012): "Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments," *Econometric Theory*, 28, 42–86.

Chao, J. C., N. R. Swanson, and T. Woutersen (2023): "Jackknife estimation of a cluster-sample IV regression model with many weak instruments," *Journal of Econometrics*, 235, 1747–1769.

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. Newey (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.

Dufour, 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*.

Hall, A. R. and A. Inoue (2003): "The large sample behaviour of the generalized method of moments estimator in misspecified models," *Journal of Econometrics*, 114, 361–394.

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.

Heckman, J. J. and E. Vytlacil (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73, 669–738.

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.

Lee, S. (2018): "A consistent variance estimator for 2SLS when instruments identify different LATEs," *Journal of Business & Economic Statistics*, 36, 400–410.

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. (2003): "A conditional likelihood ratio test for structural models," *Econometrica*, 71, 1027–1048.

——— (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.

Van de Sijpe, N. and F. Windmeijer (2023): "On the power of the conditional likelihood ratio and related tests for weak-instrument robust inference," *Journal of Econometrics*, 235, 82–104.

Yap, L. (2023): "Valid Wald Inference with Many Weak Instruments," *arXiv preprint arXiv:2311.15932*.

# A Additional Details

## A.1 Supplementary Material

**High-level Assumptions for Inference.** 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 $h_2(i,j)$ be a product of any number of $G_{i_1 i_2}, i_1 \neq i_2$, and $\check{M}_{j_1 j_2}, j_1 \neq j_2$ with $i_1, i_2, j_1, j_2 \in \{i,j\}$. Similarly, $h_3(i,j)$ denotes a product of any number of $G_{i_1 i_2}, i_1 \neq i_2$, and $\check{M}_{j_1 j_2}, j_1 \neq j_2$ with $i_1, i_2, j_1, j_2 \in \{i,j,k\}$ such that every index in $\{i,j,k\}$ occurs at least once as an index of either $G_{i_1 i_2}$ or $\check{M}_{j_1 j_2}$. Let $h_4(i,j,k,l)$ denote a product of any number of $G_{i_1 i_2}, i_1 \neq i_2$ and $\check{M}_{j_1 j_2}, j_1 \neq j_2$ with $i_1, i_2, j_1, j_2 \in \{i,j,k,l\}$ such that every index in $\{i,j,k,l\}$ occurs at least once as an index of either $G_{i_1 i_2}$ or $\check{M}_{j_1 j_2}$, and there is no partition such that $h_4(i_1,i_2,j_1,j_2) = h_2(i_1,i_2) h_2(j_1,j_2)$, where $i_1,i_2,j_1,j_2$ are all different indices. For instance, $h_4(i,j,k,l)$ could be $G_{ij}\check{M}_{ik,-il}\check{M}_{lj,-ijk}$ but not $G_{ij}\check{M}_{lk,-il}$. Let $\sum_{i\neq j}^n = \sum_i \sum_{j\neq i}$ so that sums without the $n$ superscript are still sums of individual indices, but sums 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 similarly. 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 $G'$. $1\{\cdot\}$ is an indicator function that takes the value 1 if the argument is true and 0 otherwise. $I\{\cdot\}$ is a function that takes value 1 if the argument is true and -1 if false.

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

(a) $\sum_j F_{ij}^2 \leq C$, $\sum_{j\neq k}^n \left(\sum_{i\neq j,k} G_{ij} F_{ik}\right)^2 \leq C \sum_{j\neq k}^n G_{jk}^2$, $\sum_{j\neq k}^n \left(\sum_{i\neq j,k} G_{ji} G_{ki}\right)^2 \leq C \sum_{j\neq k}^n G_{jk}^2$, 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 $\check{M}$ and $G$ cannot be too large. The conditions primarily place a restriction on the types of $G$ that can be used: for instance, a $G$ matrix that contains all 1's is excluded. Note that other elements of the covariance matrix can be analogously shown to be consistent using the same strategy by using the lemmas from Appendix B by using $\tilde{R}_{Yi} := \sum_{j\neq i} G_{ij} R_{Yj}$ instead of $\tilde{R}_{\Delta i}$ where required.

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

$$\sum_i \left(\sum_{j\neq i} h_2(i,j) R_{mj}\right)^2 = \sum_i \left(\sum_{j\in\mathcal{N}_{p(i)}\setminus\{i\}} h_2(i,j) R_{mp(i)}\right)^2 = \sum_p \sum_{i\in\mathcal{N}_p} \left(\sum_{j\in\mathcal{N}_p\setminus\{i\}} h_2(i,j) \pi_{mp}\right)^2$$

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

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

$$\sum_i \left( \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} h_4 \left(i,j,k,l\right) R_{ml} \right)^2 = \sum_p \pi_{mp}^2 \sum_{i \in \mathcal{N}_p} \left( \sum_{j \in \mathcal{N}_p \setminus \{i\}} \sum_{k \in \mathcal{N}_p \setminus \{i,j\}} \sum_{l \in \mathcal{N}_p \setminus \{i,j,k\}} h_4 \left(i,j,k,l\right) \right)^2$$
$$\leq C \sum_p \sum_{i \in \mathcal{N}_p} \pi_{mp}^2 \left(c-1\right)^2 \left(c-2\right)^2 \left(c-3\right)^2 \leq C \sum_i \tilde{R}_{mi}^2.$$

The upper bound is fairly loose because it merely counts the number of nonzero entries in $h_4$. When every judge has a large number of cases, since $h_4$ contains only entries from the projection matrix, the inner sum is still bounded and the assumption is satisfied.

**Comparing MO Variance Estimand with L3O.** The Matsushita and Otsu (2022) variance estimator presented in Equation (4) is biased in general. The model of Section 2.1 implies:

$$\begin{aligned} E\left[\hat{\Psi}_{MO}\right] &= \sum_i M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_i M_{ii}^2 R_i^2 E\left[\nu_i^2\right] + \sum_i \sum_{j \neq i} P_{ij}^2 E\left[\nu_i^2\right] E\left[\eta_j^2\right] + \sum_i \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E\left[\eta_j^2\right] \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 \left(R_i R_{\Delta i} R_j R_{\Delta j} + E\left[\eta_i \nu_i\right] R_j R_{\Delta j} + R_i R_{\Delta i} E\left[\eta_j \nu_j\right] + E\left[\eta_i \nu_i\right] E\left[\eta_j \nu_j\right]\right). \end{aligned} \tag{16}$$

As a corollary of Equation (8), when $G = P$, by observing that $P$ is symmetric, and that since $PR = I$, we have $\sum_{j \neq i} P_{ij} R_j = \sum_{j \neq i} P_{ji} R_j = M_{ii} R_i$, so

$$\begin{aligned} \text{Var}\left( \sum_i \sum_{j \neq i} P_{ij} e_i X_j \right) &= \sum_i E\left[\nu_i^2\right] M_{ii}^2 R_i^2 + \sum_i \sum_{j \neq i} P_{ij}^2 \left(E\left[\nu_i^2\right] E\left[\eta_j^2\right] + E\left[\eta_i \nu_i\right] E\left[\eta_j \nu_j\right]\right) \\ &\quad + 2 \sum_i E\left[\nu_i \eta_i\right] M_{ii}^2 R_i R_{\Delta i} + \sum_i E\left[\eta_i^2\right] M_{ii}^2 R_{\Delta i}^2. \end{aligned} \tag{17}$$

If the $R_\Delta$'s are zero, then $\hat{\Psi}_{MO}$ is unbiased. Nonetheless, 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 $\frac{1}{\sqrt{K}} \sum_i M_{ii} R_i^2 \to C_S < \infty$ and $\frac{1}{\sqrt{K}} \sum_i M_{ii} R_{\Delta i}^2 \to C < \infty$ with weak identification and weak heterogeneity, then the excess terms in $\frac{1}{K} E\left[\hat{\Psi}_{MO}\right]$ 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.

In general, heterogeneity does not make the MO variance estimator any more or less conservative than L3O. In the simple case with judge instruments and $G = P$, we have:

$$\begin{aligned} E\left[\hat{\Psi}_{MO}\right] - Var\left( \sum_i \sum_{j \neq i} P_{ij} e_i X_j \right) &= \sum_i M_{ii}^2 R_i^2 R_{\Delta i}^2 + \sum_i \sum_{j \neq i} P_{ij}^2 R_i R_{\Delta i} R_j R_{\Delta j} \\ &\quad + \sum_i \sum_{j \neq i} P_{ij}^2 R_{\Delta i}^2 E\left[\eta_j^2\right] - \sum_i M_{ii}^2 R_{\Delta i}^2 E\left[\eta_i^2\right] + 2 \sum_i \sum_{j \neq i} P_{ij}^2 E\left[\eta_i \nu_i\right] R_j R_{\Delta j} - 2 \sum_i E\left[\nu_i \eta_i\right] M_{ii}^2 R_i R_{\Delta i} \\ &= \sum_i M_{ii} R_{\Delta i}^2 R_i^2 - \sum_i M_{ii} (1 - 2P_{ii}) R_{\Delta i}^2 E\left[\eta_i^2\right] - 2 \sum_i M_{ii} (1 - 2P_{ii}) E\left[\eta_i \nu_i\right] R_i R_{\Delta i} \\ &= \sum_i M_{ii} R_{\Delta i}^2 \left(R_i^2 - (1 - 2P_{ii}) E\left[\eta_i^2\right]\right) - 2 \sum_i M_{ii} (1 - 2P_{ii}) E\left[\eta_i \nu_i\right] R_i R_{\Delta i} \end{aligned}$$

which can be positive or negative. The second equality uses the fact that $P$ and $M$ are non-zero only for observations that share the same judge, and when that occurs, they have the same $R, R_Y, E[\eta_i^2]$, and $E[\zeta_i^2]$, and that $\sum_{j \neq i} P_{ij}^2 = P_{ii} M_{ii}$.

To compare the confidence sets of MO and L3O, observe that the shape of the confidence set depends on the coefficient on $\beta_0^2$. In particular, for $\Psi_2 := \sum_i \left(\sum_{j\neq i} P_{ij}X_j\right)^2 X_i^2 + \sum_i \sum_{j\neq i} P_{ij}^2 X_i^2 X_j^2$, MO is unbounded when $\left(\sum_{i\neq j}^n P_{ij}X_iX_j\right)^2 - q\Psi_2 < 0$ and L3O is unbounded when $\left(\sum_{i\neq j}^n P_{ij}X_iX_j\right)^2 - qB_2 < 0$. In the simple judges case without covariates, the expected coefficients can be compared. With

$$E\left[\Psi_2\right] = \sum_i \left(\left(\sum_{j\neq i} P_{ij}R_j\right)^2 + \sum_{j\neq i} P_{ij}^2 E\left[\eta_j^2\right]\right) E[X_i^2] + \sum_i \sum_{j\neq i} P_{ij}^2 E[X_i^2]E[X_j^2],$$

the difference is:

$$\begin{aligned} E\left[\Psi_2\right] - E\left[B_2\right] &= \sum_i \left(\sum_{j\neq i} P_{ij}R_j\right)^2 R_i^2 + \sum_i \sum_{j\neq i} P_{ij}^2 \left(R_i^2R_j^2 + 3E\left[\eta_i^2\right]R_j^2\right) - 3\sum_i \left(\sum_{j\neq i} P_{ij}R_j\right)^2 E\left[\eta_i^2\right] \\ &= \sum_i M_{ii}R_i^2 \left(R_i^2 - 3\left(1-2P_{ii}\right)E\left[\eta_i^2\right]\right), \end{aligned}$$

where the second equality uses the same trick as before.

**Restrictions on $\mu$.** The restrictions on $\mu$ in Equation (14) can be extended to the case with a saturated covariate and instrument space. To be precise, saturation is defined in Section 2 of Evdokimov and Kolesár (2018). All individuals can be partitioned into $L$ covariate groups, so with group index $G_i \in \{1,\cdots,L\}$, we have covariates $W_{i,l} = 1\{G_i = l\}$. We also have an instrument $S_i$ that takes $M+1$ possible values in each group, and these values for every group $l$ are labelled $s_{l0},\cdots,s_{lM}$. Then, the vector of instruments has dimension $K = ML$ and $Z_{i,lm} = 1\{S_i = s_{lm}\}$. Adapting Equation (14) to the case with covariates,

$$\mu = \left(\frac{1}{\sqrt{K}}\sum_i\sum_{j\neq i} G_{ij}R_{Yi}R_{Yj}, \frac{1}{\sqrt{K}}\sum_i\sum_{j\neq i} G_{ij}R_{Yi}R_j, \frac{1}{\sqrt{K}}\sum_i\sum_{j\neq i} G_{ij}R_iR_j\right)'.$$

**Proposition 4.** *If $\mu$ is defined such that $G = \left(I - \operatorname{diag}\left(H_Q\right)\right)^{-1} H_Q - \left(I - \operatorname{diag}\left(H_W\right)\right)^{-1} H_W$, and the regression is fully saturated, then $\mu_1 \geq 0, \mu_3 \geq 0$ and $\mu_2^2 \leq \mu_1\mu_3$.*

## A.2 Details for Section 2

**Lemma 3.** *Consider the model of Section 2. Suppose $h \neq 0$ and $Ks^2 > 0$. Then, $E\left[T_{ee}\right] \neq 0$ for all real $\beta_0$.*

**Data Generating Process.** Data is generated from an environment with $E[\varepsilon_i] = 0$, and $\int_0^1 f(v)dv = \beta$. To run a regression on judge indicators (without an intercept) in the reduced-form system, I make a transformation $\check{X} = 2X - 1$ so that the reduced-form equations can be written as:

$$\check{X}_i = Z_i'\pi + \eta_i, \text{ and } Y_i = Z_i'\pi_Y + \zeta_i,$$

so $\pi_k = \pi_{Yk} = 0$ for the base judge. The reduced-form errors are: $\eta_i = I\left\{\lambda_{k(i)} - v_i \geq 0\right\} - \pi_{k(i)}$ and $\zeta_i = 1\left\{\lambda_{k(i)} - v_i \geq 0\right\} f\left(v_i\right) + \varepsilon_i - \pi_{Yk(i)}$ respectively. With $\pi_{\Delta k} = \pi_{Yk} - \pi_k\beta$, the reduced-form parameters for the groups of judges are derived in Table 7. Since the coefficient of the base judge is normalized to zero, the implementation without covariates in simulations excludes the intercept and uses indicators for all judges, instead of omitting the base judge and having an intercept. This implementation results in a block diagonal projection matrix, which aids computational speed, while retaining the interpretation of $\pi$'s in the

Table 7: Parameters for Simple Example

| $\lambda_k$ | $\frac{1}{2}-s$ | $\frac{1}{2}-\frac{1}{2}s$ | $\frac{1}{2}$ | $\frac{1}{2}+\frac{1}{2}s$ | $\frac{1}{2}+s$ |
|---|---|---|---|---|---|
| $\beta_k$ | $\beta-\frac{h}{s}$ | $\beta+2\frac{h}{s}$ | NA | $\beta-2\frac{h}{s}$ | $\beta+\frac{h}{s}$ |
| $\pi_k$ | $-s$ | $-\frac{1}{2}s$ | 0 | $\frac{1}{2}s$ | $s$ |
| $\pi_{Yk}$ | $-s\beta+h$ | $-\frac{1}{2}s\beta-h$ | 0 | $\frac{1}{2}s\beta-h$ | $s\beta+h$ |
| $\pi_{\Delta k}$ | $h$ | $-h$ | 0 | $-h$ | $h$ |

reduced-form model. The $f(v)$ that delivers the parameters in Table 7 is

$$f(v)=\begin{cases}-s\beta+h & v\in[0,\frac{1}{2}-s]\\ \frac{1}{s}(1-s)\left(-\frac{1}{2}s\beta-h\right)-\frac{1}{s}(1-2s)(-s\beta+h) & v\in(\frac{1}{2}-s,\frac{1}{2}-\frac{1}{2}s]\\ \frac{1}{s}(1-s)\left(\frac{1}{2}s\beta+h\right) & v\in(\frac{1}{2}-\frac{1}{2}s,\frac{1}{2}]\\ \frac{1}{s}(1+s)\left(\frac{1}{2}s\beta-h\right) & v\in(\frac{1}{2},\frac{1}{2}+\frac{1}{2}s]\\ \frac{1}{s}(1+2s)(s\beta+h)-\frac{1}{s}(1+s)\left(\frac{1}{2}s\beta-h\right) & v\in(\frac{1}{2}+\frac{1}{2}s,\frac{1}{2}+s]\\ \frac{\beta-\left(\frac{1}{2}+s\right)(s\beta+h)}{\frac{1}{2}-s} & v\in(\frac{1}{2}+s,1]\end{cases}. \tag{18}$$

To generate the data in the simulation, I draw $v_i\sim U[0,1]$ as implied by the structural model, then generate $\zeta_i\mid v_i\sim N(\sigma_{\varepsilon v}v_i,\sigma_{\varepsilon\varepsilon})$. Hence, $\sigma_{\varepsilon v}$ and $\sigma_{\varepsilon\varepsilon}$ control the correlation between $\eta_i$ and $\zeta_i$, with $\sigma_{\varepsilon\varepsilon}=0$ corresponding to perfect correlation. In the base case, I set $\sigma_{\varepsilon\varepsilon}=0.1$ and $\sigma_{\varepsilon v}=0.3$. With the given $\pi_k,\pi_{Yk}$, the observable variables are generated from $\check{X}_i=I\{\pi_{k(i)}>v_i\}$ and $Y_i=\pi_{Yk(i)}+\zeta_i$.

**Derivations for Constructed Instrument** Using the notation for the just-identified IV AR test in Section 2.4,

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

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

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

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

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

**Clustering by Judges**. If we just use the JIVE t-ratio with clustering, weak identification is still a problem, and we should similarly get over-rejection. The just-identified AR with clustered standard errors uses the following estimator:

$$\hat{V}_{clus} = \frac{\sum_i \sum_{j\in\mathcal{N}_i} \tilde{X}_i \hat{\varepsilon}_i \tilde{X}_j \hat{\varepsilon}_j}{\left(\sum_i \tilde{X}_i^2\right)^2},$$

where $\mathcal{N}_i$ is the neighborhood of $i$ (i.e., the set of observations that share the same cluster as $i$). Expanding $\hat{V}_{clus}$ using the same steps as before,

$$\hat{V}_{clus} = \frac{\sum_i \sum_{j\in\mathcal{N}_i} \tilde{X}_i \tilde{X}_j \left(e_i e_j \left(\sum_k \tilde{X}_k^2\right)^2 - 2\tilde{X}_i e_j \left(\sum_k e_k \tilde{X}_k\right)\left(\sum_k \tilde{X}_k^2\right) + \tilde{X}_i \tilde{X}_j \left(\sum_k e_k \tilde{X}_k\right)^2\right)}{\left(\sum_i \tilde{X}_i^2\right)^4}.$$

Using the fact that $\frac{1}{n}\sum_k e_k \tilde{X}_k = o_P(1)$, the dominant term is: $\sum_i \sum_{j\in\mathcal{N}_i} \tilde{X}_i \tilde{X}_j e_i e_j$, which is analogous to the previous derivation. The expansion steps are analogous to that required to derive $V_{LM}$, so they are omitted. Then,

$$E\left[\sum_i \sum_{j\in\mathcal{N}_i} \tilde{X}_i \tilde{X}_j e_i e_j\right] = E\left[\sum_i \sum_{j\in\mathcal{N}_i} \left(\sum_{k\neq i} P_{ik}\left(R_k+\eta_k\right)\right)\left(\sum_{k\neq j} P_{jk}\left(R_k+\eta_k\right)\right)\left(R_{\Delta i}+\nu_i\right)\left(R_{\Delta j}+\nu_j\right)\right]$$
$$= E\left[\sum_i \sum_{j\in\mathcal{N}_i} \left(M_{ii}R_i + \sum_{k\neq i} P_{ik}\eta_k\right)\left(M_{jj}R_j + \sum_{k\neq j} P_{jk}\eta_k\right)\left(R_{\Delta i}R_{\Delta j} + \nu_i R_{\Delta j} + R_{\Delta i}\nu_j + \nu_i\nu_j\right)\right]$$
$$= \sum_i \sum_{j\in\mathcal{N}_i} \left(M_{ii}M_{jj}R_i R_j R_{\Delta i} R_{\Delta j} + R_{\Delta i} R_{\Delta j} \sum_{k\neq i,j} P_{ik} P_{jk} E\left[\eta_k^2\right]\right)$$
$$+ 2\sum_i \sum_{j\in\mathcal{N}_i} M_{ii} R_i R_{\Delta j} P_{ji} E\left[\eta_i \nu_i\right] + \sum_i M_{ii}^2 R_i^2 E\left[\nu_i^2\right]$$
$$+ \sum_i \sum_{k\neq i} P_{ik}^2 E\left[\nu_i^2\right] E\left[\eta_k^2\right] + \sum_i \sum_{j\in\mathcal{N}_i\setminus\{i\}} P_{ij}^2 E\left[\nu_i \eta_i\right] E\left[\nu_j \eta_j\right].$$

By applying the fact that the entries of the projection matrix are nonzero only when the observations share the same judge, the expression simplifies further:

$$E\left[\sum_i \sum_{j\in\mathcal{N}_i} \tilde{X}_i \tilde{X}_j e_i e_j\right] = \sum_i \sum_{j\in\mathcal{N}_i} M_{ii}M_{jj}R_i R_j R_{\Delta i} R_{\Delta j} + \sum_i M_{ii}^2 R_{\Delta i}^2 E\left[\eta_i^2\right]$$
$$+ 2\sum_i M_{ii} R_i R_{\Delta i} E\left[\eta_i \nu_i\right] + \sum_i M_{ii}^2 R_i^2 E\left[\nu_i^2\right] + \sum_i \sum_{j\neq i} P_{ij}^2 \left(E\left[\nu_i^2\right] E\left[\eta_j^2\right] + E\left[\nu_i \eta_i\right] E\left[\nu_j \eta_j\right]\right).$$

Compared to the true variance in Equation (17), due to the own-observation bias, we have an extra $\sum_i \sum_{j\in\mathcal{N}_i} M_{ii}M_{jj}R_i R_j R_{\Delta i} R_{\Delta j}$ term, and the estimand here has $\sum_i M_{ii} R_i R_{\Delta i} E\left[\eta_i \nu_i\right]$ instead of $\sum_i M_{ii}^2 R_i R_{\Delta i} E\left[\eta_i \nu_i\right]$. Even though $\sum_i \sum_{j\in\mathcal{N}_i} M_{ii}M_{jj}R_i R_j R_{\Delta i} R_{\Delta j} \geq 0$, $\sum_i M_{ii}(1-M_{ii}) R_i R_{\Delta i} E\left[\eta_i \nu_i\right]$ could be positive or negative, so the clustered variance estimand could either over or underestimate the true variance.

## A.3 Details for Section 4

### A.3.1 Existence of Structural Model

This section presents a structural model, then argues that any reduced-form model in the form of Equation (14) can be justified by this structural model.

**Example 1.** Consider a linear potential outcomes model with an instrument $Z$ that is a vector of indicators for judges, each with $c = 5$ cases, a continuous endogenous variable $X$, and outcome $Y$:

$$X_i(z) = z'\pi + v_i, \qquad Y_i(x) = x(\beta + \xi_i) + \varepsilon_i, \text{ and}$$
$$\begin{pmatrix} \varepsilon_i \\ \xi_i \\ v_i \end{pmatrix} \mid k(i) = k \sim N\left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\varepsilon\varepsilon} & \sigma_{\varepsilon\xi} & \sigma_{\varepsilon v} \\ \cdot & \sigma_{\xi\xi} & \sigma_{\xi vk} \\ \cdot & \cdot & \sigma_{vv} \end{pmatrix} \right). \tag{19}$$

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 $C_S = \frac{1}{\sqrt{K}} \sum_k (c-1)\pi_k^2$. The $\pi_k$'s are constructed as such: with $s = \sqrt{C_S/\sqrt{K}/(c-1)}$, set $\pi_k = 0$ for the base judge, $\pi_k = -s$ for half the judges and $\pi_k = s$ for the other half. The heterogeneity covariances $\sigma_{\xi vk}$ are constructed so that $\sum_k \pi_k = 0, \sum_k \sigma_{\xi vk} = 0$, and $\sum_k \pi_k \sigma_{\xi vk} = 0$. With $C_H$ characterizing the heterogeneity in the model, and $h = \sqrt{C_H/\sqrt{K}/(c-1)}$, set $\sigma_{\xi vk} = 0$ of the base judge; among judges with $\pi_k = s$, half of them have $\sigma_{\xi vk} = h$ and the other half $\sigma_{\xi vk} = -h$. The same construction of $\sigma_{\xi vk}$ applies for judges with $\pi_k = -s$.

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, $\varepsilon_i$ as the noise in the intercept of the outcome equation, and $\xi_i$ as the individual-level treatment effect heterogeneity. Further, $\sigma_{\xi vk}$ characterizes the extent of treatment effect heterogeneity. The observed outcome in a model with constant treatment effects is $Y_i(X_i) = X_i\beta + \breve{\varepsilon}_i$, with $E[\breve{\varepsilon}_i]$=0. When $\sigma_{\xi vk} = 0$, regardless of the values of $\sigma_{\varepsilon\xi}, \sigma_{\xi\xi}$, the observed outcome of Equation (19) can be written as $Y_i(X_i) = X_i\beta + \breve{\varepsilon}_i$ where $E[\breve{\varepsilon}_i] = E[X_i\xi_i + \varepsilon_i] = E[X_i E[\xi_i \mid X_i]] = 0$, which resembles the constant treatment effect case.

**Lemma 4.** *Consider the model of Example 1. If $\sqrt{K}s^2 \to \tilde{C}_S < \infty$ and $\sqrt{K}h^2 \to \tilde{C}_H < \infty$, then*

$$\sigma_{11} = \frac{4}{\sigma_{33}} \left( \sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right)^2 + o(1), \quad \sigma_{12} = 2\frac{\sigma_{23}}{\sigma_{33}} \left( \sigma_{22} - \frac{\sigma_{23}^2}{2\sigma_{33}} \right) + o(1), \quad \sigma_{13} = \frac{\sigma_{23}^2}{\sigma_{33}} + o(1),$$
$$\sigma_{22} = \frac{c-1}{c} \left( \sigma_{vv} \left( \sigma_{\varepsilon\varepsilon} + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\varepsilon v}\beta \right) + (\sigma_{vv}\beta + \sigma_{\varepsilon v})^2 \right) + o(1),$$
$$\sigma_{33} = 2\frac{c-1}{c}\sigma_{vv}^2 + o(1), \quad \sigma_{23} = 2\frac{c-1}{c}\sigma_{vv} (\sigma_{vv}\beta + \sigma_{\varepsilon v}) + o(1), \text{ and}$$
$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \sqrt{K}(c-1)(s^2\beta^2 + h^2) \\ \sqrt{K}(c-1)s^2\beta \\ \sqrt{K}(c-1)s^2 \end{pmatrix} = (c-1) \begin{pmatrix} C_S\beta^2 + C_H \\ C_S\beta \\ C_S \end{pmatrix}.$$

**Proposition 5.** *In the model of Example 1 with $\sqrt{K}s^2 \to \tilde{C}_S < \infty$ and $\sqrt{K}h^2 \to \tilde{C}_H < \infty$, for any $\sigma_{22}, \sigma_{23}, \sigma_{33}$ such that $\sigma_{22}, \sigma_{33} > 0$, $\sigma_{23}^2 \leq \sigma_{22}\sigma_{33}$ and $\mu$ such that $\mu_1 \geq 0, \mu_3 > 0$, $\mu_2^2 \leq \mu_1\mu_3$, the following values of structural parameters:*

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

$$\sigma_{vv} = \sqrt{\frac{\sigma_{33} c}{2(c-1)}}, \ \textit{and} \quad \sigma_{\varepsilon v} = \frac{1}{\sigma_{vv}} \left( \frac{\sigma_{23} c}{2(c-1)} - \sigma_{vv}^2 \beta \right),$$

*satisfy the equations in Lemma 4, and* $\det(\Sigma_{SF})/h \to C_D \geq 0$.

Due to Proposition 5, since the principal submatrices of $\Sigma_{SF}$ are positive semidefinite asymptotically, $\Sigma_{SF}$ is a symmetric positive semidefinite matrix asymptotically. The proposition thus implies that when the $\sigma$'s and $\mu$ satisfy the conditions, there exists structural parameters that can generate the given $\mu$ and $\Sigma$ asymptotically. Hence, there are no further restrictions on $\mu$ from the observed $\Sigma$ in Example 1.

### A.3.2 Numerical Results for Power

Beyond the theoretical optimality results of Section 4, this section presents numerical results for power in environments not covered by the theory. I first consider one-sided tests beyond the set $\mathcal{S}$ covered by the theory, then weighted average power for two-sided tests rather than the class of unbiased tests.

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) (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 $\Sigma$ matrix of the form:

$$\Sigma = \begin{pmatrix} 2 & 2\rho & 2\rho^2 \\ \cdot & 1+\rho^2 & 2\rho \\ \cdot & \cdot & 2 \end{pmatrix}, \tag{20}$$

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

In the numerical exercises, I display the rejection rate across 500 independent draws from $X^* \sim N(\mu, \Sigma)$ at each point on the $\mu_2$ axis, across several $\mu_1, \mu_3$ values 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.

Figure 2 uses a one-sided LM test, with a large covariance at $\rho = 0.9$. 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 2, when $\mu_2$ is small enough for $\mu_1 = 1, \mu_3 = 4$, LM achieves the power envelope, but as $\mu_2$ 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.

In Figure 3, $\Sigma$ is calibrated by using the $\Sigma$ matrix calculated from the Angrist and Krueger (1991) application, so after appropriate normalizations, $\rho = 0.37$. 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 application considered, the gains are small.

Instead of considering a point alternative, we may be more interested in testing against a composite alternative. Here, 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 inequality constraints. Figures 4 and 5 present 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 that substantially outperform LM, this result is compatible with Proposition 3. EMW is a biased test as there are points in the alternative space that are not a part of the grid where LM outperforms EMW. Nonetheless, Figure 5 suggests that, when using the empirical covariance, LM does not perform substantially worse than EMW.

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

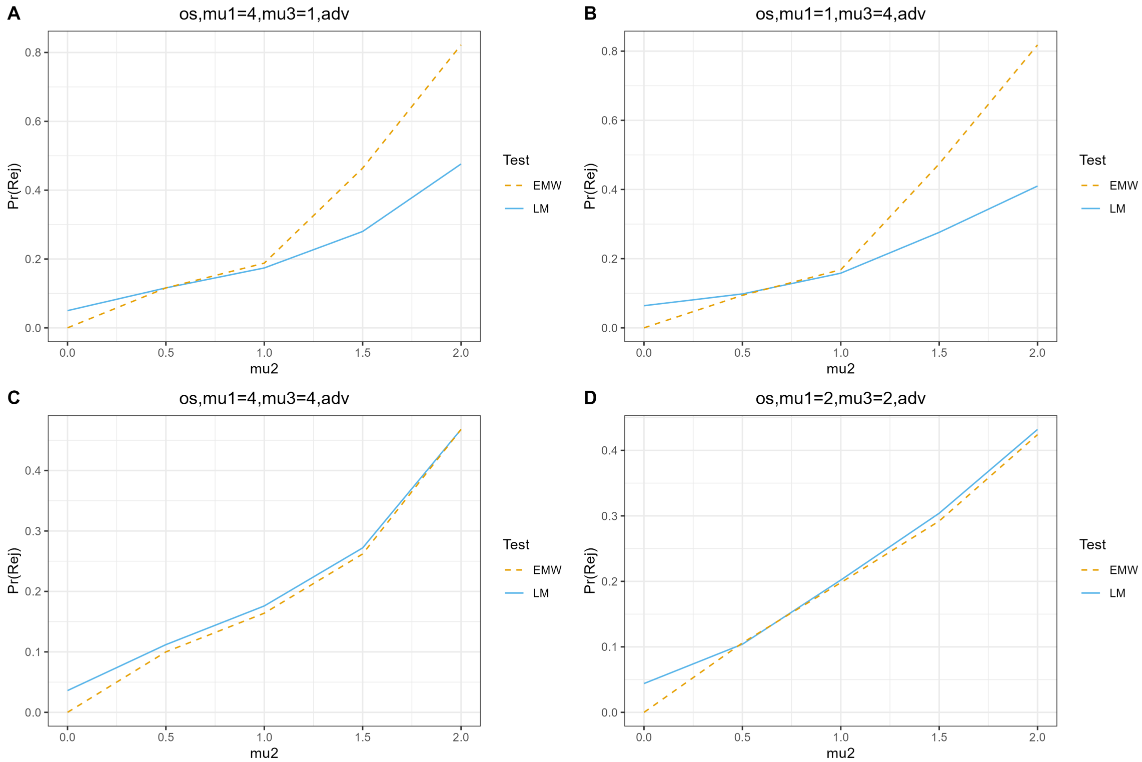



Figure 3: One-sided test with $\rho = 0.37$

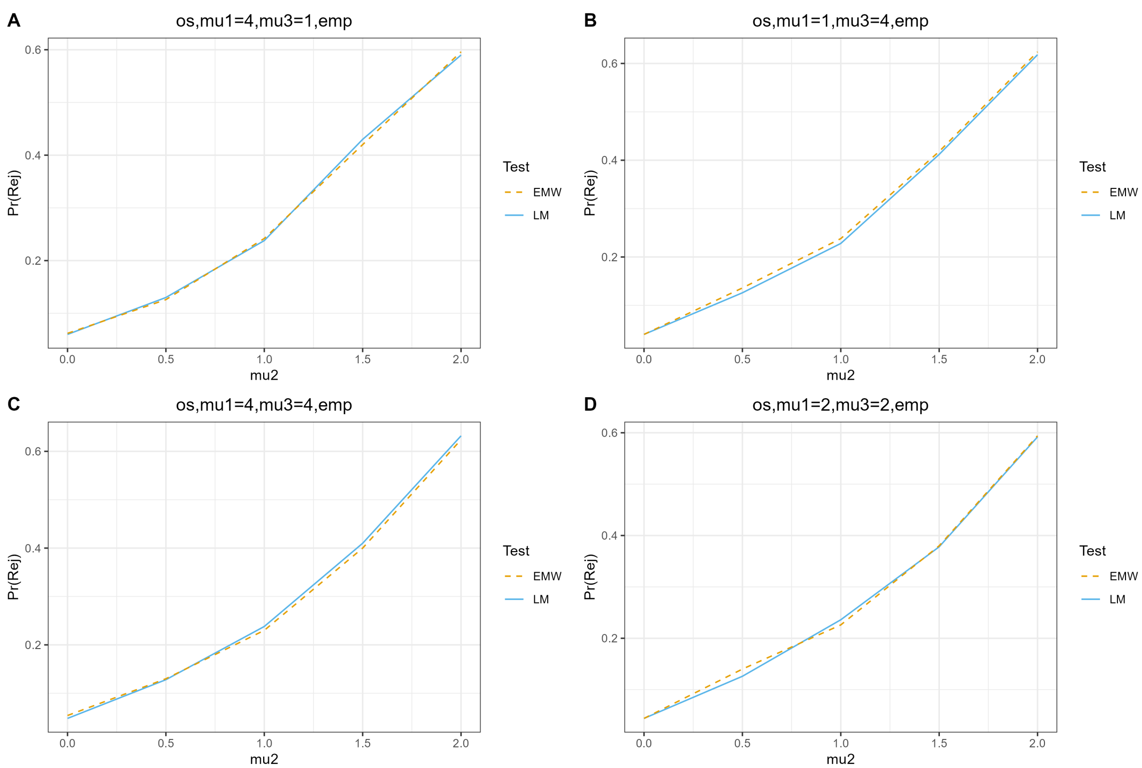


Figure 4: Uniform Weighting on grid of alternatives with $\rho = 0.9$

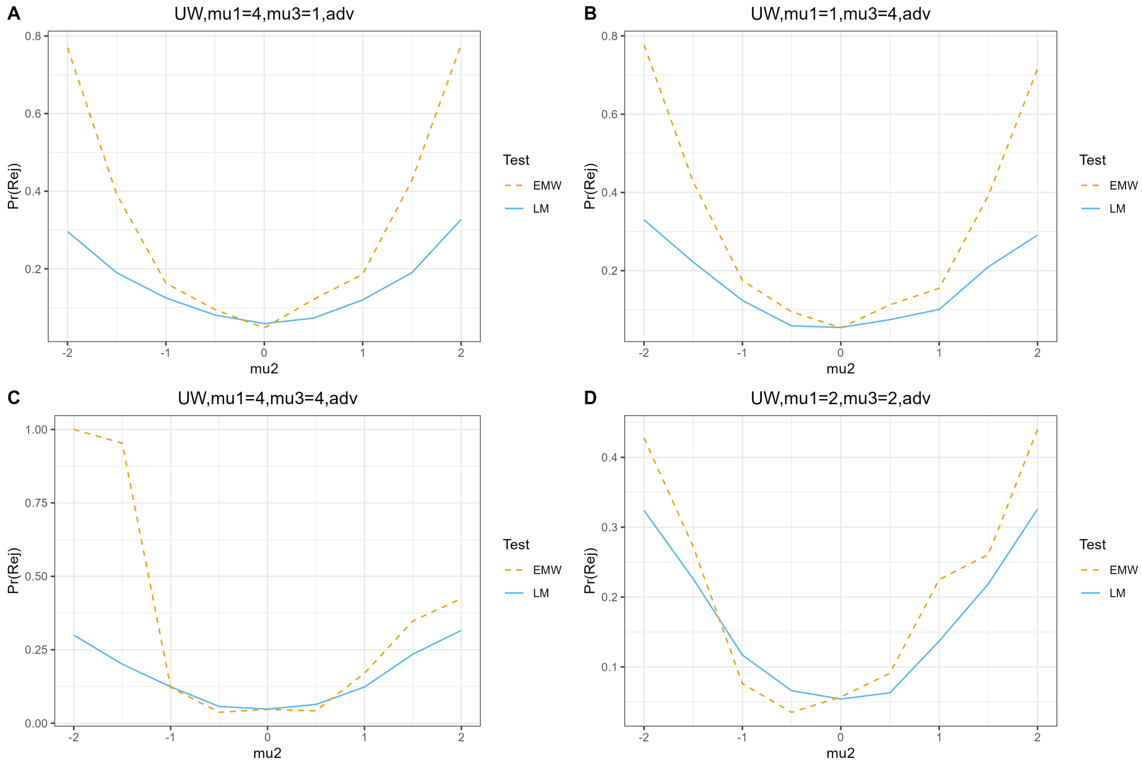



Figure 5: Uniform Weighting on grid of alternatives with $\rho = 0.37$

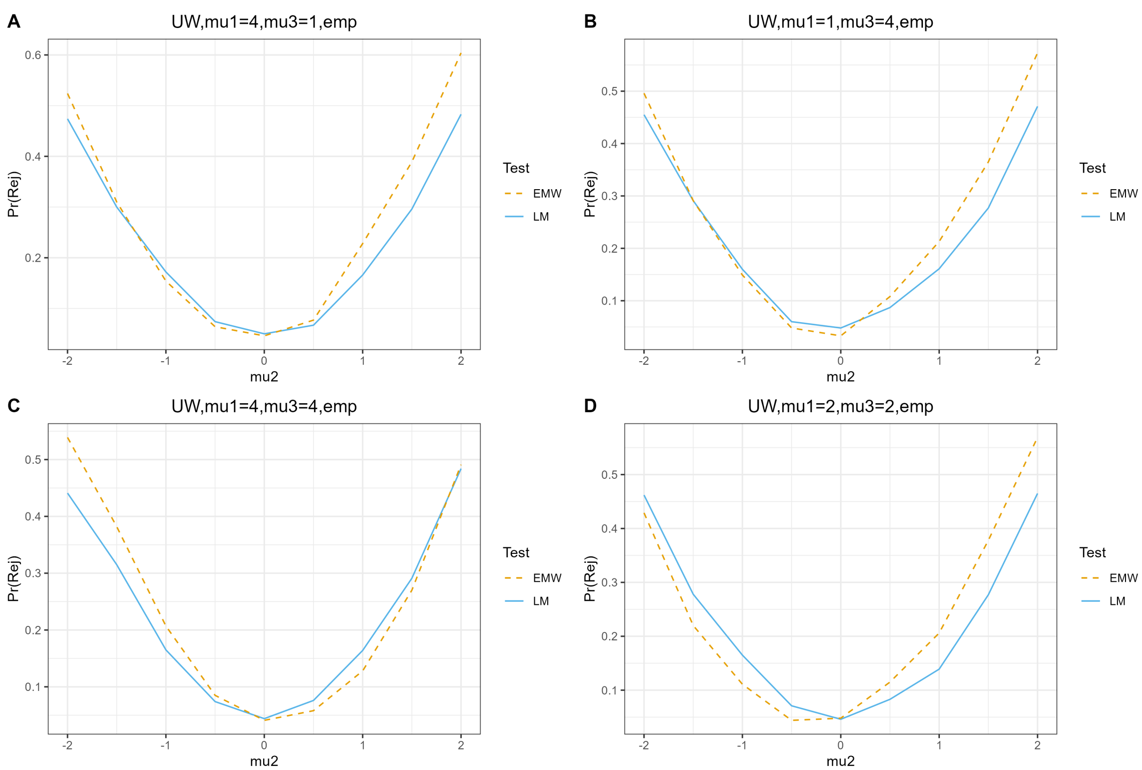


Table 8: Rejection rates under the null for nominal size 0.05 test for continuous $X$

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = C_S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$ | 0.061 | 0.017 | 1.000 | 0.061 | 0.079 | 0.078 | 0.042 | 0.044 |
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 0.952 | 0.022 | 1.000 | 0.073 | 0.087 | 0.084 | 0.058 | 0.055 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 1.000 | 0.009 | 1.000 | 0.096 | 0.076 | 0.127 | 0.053 | 0.050 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 1.000 | 0.006 | 1.000 | 0.103 | 0.061 | 0.127 | 0.059 | 0.052 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 0.986 | 0.033 | 0.109 | 0.057 | 0.062 | 0.064 | 0.056 | 0.047 |
| $C_H = 3, C_S = 3$ | 1.000 | 0.036 | 0.168 | 0.055 | 0.078 | 0.087 | 0.055 | 0.047 |
| $C_H = 3, C_S = 0$ | 1.000 | 0.048 | 0.184 | 0.058 | 0.106 | 0.088 | 0.053 | 0.057 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 1.000 | 0.089 | 0.049 | 0.063 | 0.083 | 0.080 | 0.061 | 0.058 |
| $C_H = 0, C_S = 2$ | 1.000 | 0.207 | 0.045 | 0.054 | 0.243 | 0.135 | 0.057 | 0.045 |
| $C_H = 0, C_S = 0$ | 1.000 | 0.337 | 0.051 | 0.042 | 0.413 | 0.127 | 0.045 | 0.048 |
| $C_H = C_S = 0, \sigma_{\varepsilon v} = 1$ | 1.000 | 1.000 | 0.044 | 0.042 | 1.000 | 0.157 | 0.052 | 0.044 |

Notes: Data generating process corresponds to Example 1. 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 $\sigma_{\xi\xi} = \sigma_{\varepsilon\xi} = \sigma_{\xi v} = 0$, corresponding to constant treatment effects. Procedures are described in Table 2.

## A.4 Further Simulation Results

This section reports simulation results from several structural models to assess how well various procedures control for size. Since the nominal size is 0.05, and data is generated under the null, the target rejection rate is 0.05. Across the board, the L3O method performs well, and for all existing procedures, there exists at least one design where they perform badly. Comments for the procedures are in Table 2.

### A.4.1 Continuous Treatment

This subsection reports results for a simulation based on Example 1 that has a continuous $X$. Table 8 reports results with $K = 500$ and Table 9 reports results for $K = 40$. The L3O rejection rates are closer to the nominal rate than the existing procedures in the literature, albeit worse with a smaller $K$. MS has high rejection rates with strong heterogeneity and EK has high rejection rates with weak instruments. Notably, with perfect correlation and an irrelevant instrument, EK can achieve 100% rejection in the simulation with $K = 500$. The procedures that use the LM statistic are MO, $\tilde{X}$-AR, L3O and LMorc; they differ only in their variance estimation. Hence, while $\tilde{X}$-AR and MO over-reject, the extent of over-rejection is smaller than MS and EK in the adversarial cases.

### A.4.2 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 9: Rejection Rates under the null for nominal size 0.05 test for Continuous $X$ with $K = 40$

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = C_S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$ | 0.072 | 0.022 | 0.525 | 0.051 | 0.074 | 0.068 | 0.039 | 0.055 |
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 0.238 | 0.034 | 0.388 | 0.051 | 0.074 | 0.077 | 0.055 | 0.062 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 0.547 | 0.033 | 0.475 | 0.083 | 0.096 | 0.133 | 0.077 | 0.053 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 0.651 | 0.013 | 0.511 | 0.072 | 0.088 | 0.102 | 0.068 | 0.054 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 0.213 | 0.025 | 0.109 | 0.048 | 0.057 | 0.063 | 0.055 | 0.046 |
| $C_H = 3, C_S = 3$ | 0.658 | 0.032 | 0.129 | 0.045 | 0.074 | 0.063 | 0.064 | 0.055 |
| $C_H = 3, C_S = 0$ | 0.849 | 0.049 | 0.127 | 0.063 | 0.109 | 0.103 | 0.087 | 0.057 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 0.853 | 0.105 | 0.049 | 0.064 | 0.068 | 0.098 | 0.085 | 0.056 |
| $C_H = 0, C_S = 2$ | 0.999 | 0.152 | 0.048 | 0.045 | 0.201 | 0.132 | 0.098 | 0.037 |
| $C_H = 0, C_S = 0$ | 1.000 | 0.342 | 0.052 | 0.051 | 0.439 | 0.143 | 0.080 | 0.049 |
| $C_H = C_S = 0, \sigma_{\varepsilon v} = 1$ | 1.000 | 1.000 | 0.045 | 0.040 | 1.000 | 0.179 | 0.082 | 0.045 |

Note: Designs are identical to Table 8, but $K = 40$ here.

Our unobservables are generated as follows. Draw $v_i \sim U[-1, 1]$, then generate residuals from:

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

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

The process for determining $s, h$ and $\pi_k \in \{0, -s, s\}, \sigma_{\xi vk} \in \{0, -h, h\}$ are identical to Example 1, as $s$ controls the strength of the instrument, $h$ the extent of heterogeneity, and $\beta$ 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 $\beta$ in Appendix E. A simulation is run with $K = 100$, so the sample size is smaller than the normal experiment in Example 1.

Results are presented in Table 10, 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.

### A.4.3 Incorporating Covariates

This section presents a data-generating process that involves covariates. Instead of judges, consider a model where there are $K$ states. Let $t = 1, \cdots, K$ 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

Table 10: Rejection Rates under the null for nominal size 0.05 test for binary $X$

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = C_S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$ | 0.046 | 0.049 | 0.059 | 0.045 | 0.045 | 0.045 | 0.049 | 0.054 |
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 0.097 | 0.047 | 0.177 | 0.037 | 0.038 | 0.041 | 0.051 | 0.052 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 0.727 | 0.059 | 1.000 | 0.127 | 0.051 | 0.143 | 0.058 | 0.051 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 0.891 | 0.037 | 1.000 | 0.204 | 0.067 | 0.247 | 0.059 | 0.045 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 0.092 | 0.060 | 0.051 | 0.055 | 0.057 | 0.056 | 0.055 | 0.047 |
| $C_H = 3, C_S = 3$ | 0.996 | 0.089 | 0.888 | 0.059 | 0.086 | 0.096 | 0.055 | 0.048 |
| $C_H = 3, C_S = 0$ | 1.000 | 0.124 | 0.999 | 0.101 | 0.289 | 0.181 | 0.068 | 0.052 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 0.408 | 0.058 | 0.055 | 0.043 | 0.046 | 0.046 | 0.045 | 0.041 |
| $C_H = 0, C_S = 2$ | 1.000 | 0.212 | 0.052 | 0.061 | 0.188 | 0.108 | 0.078 | 0.057 |
| $C_H = 0, C_S = 0$ | 1.000 | 0.654 | 0.046 | 0.034 | 0.750 | 0.149 | 0.069 | 0.039 |
| $C_H = C_S = 0, \sigma_{\varepsilon\varepsilon} = 0$ | 1.000 | 1.000 | 0.053 | 0.057 | 1.000 | 0.173 | 0.076 | 0.053 |

Note: The data generating process corresponds to Appendix A.4.2. 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.

Table 11: Rejection Rates under the null for nominal size 0.05 test for binary $X$ with covariates

| | TSLS | EK | $T_{ee}$(MS) | $T_{eX}$(MO) | $\tilde{X}$-t | $\tilde{X}$-AR | L3O | LMorc |
|---|---|---|---|---|---|---|---|---|
| $C_H = C_S = 3\sqrt{K}, \sigma_{\varepsilon v} = 0$ | 0.048 | 0.123 | 0.049 | 0.052 | 0.047 | 0.055 | 0.054 | 0.060 |
| $C_H = 2\sqrt{K}, C_S = 2\sqrt{K}$ | 0.072 | 0.111 | 0.052 | 0.044 | 0.041 | 0.046 | 0.050 | 0.053 |
| $C_H = 2\sqrt{K}, C_S = 2$ | 0.171 | 0.016 | 0.471 | 0.083 | 0.012 | 0.092 | 0.060 | 0.050 |
| $C_H = 2\sqrt{K}, C_S = 0$ | 0.259 | 0.002 | 0.960 | 0.126 | 0.008 | 0.135 | 0.047 | 0.058 |
| $C_H = 3, C_S = 3\sqrt{K}$ | 0.065 | 0.132 | 0.048 | 0.053 | 0.056 | 0.054 | 0.060 | 0.049 |
| $C_H = 3, C_S = 3$ | 0.131 | 0.015 | 0.108 | 0.040 | 0.003 | 0.042 | 0.044 | 0.050 |
| $C_H = 3, C_S = 0$ | 0.247 | 0.003 | 0.300 | 0.087 | 0.004 | 0.091 | 0.062 | 0.053 |
| $C_H = 0, C_S = 2\sqrt{K}$ | 0.084 | 0.099 | 0.054 | 0.041 | 0.036 | 0.043 | 0.048 | 0.050 |
| $C_H = 0, C_S = 2$ | 0.178 | 0.006 | 0.058 | 0.043 | 0.002 | 0.044 | 0.052 | 0.051 |
| $C_H = 0, C_S = 0$ | 0.246 | 0.006 | 0.048 | 0.063 | 0.005 | 0.069 | 0.081 | 0.050 |
| $C_H = C_S = 0, \sigma_{\varepsilon\varepsilon} = 0$ | 1.000 | 0.497 | 0.042 | 0.013 | 0.147 | 0.049 | 0.092 | 0.035 |

Note: The data generating process corresponds to Appendix A.4.3. 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 base group, which constitutes half the observations. For $k \neq 0$, $\pi_k$ is the coefficient for observations from state $t = k$ and have $B = 1$, and $\sigma_{\xi v k}$ 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 $\beta$, and the proof is in Appendix E. Table 11 reports the associated simulation results, which are qualitatively similar to the results described before.

# B Proofs for Section 3

## B.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 $i = 1, \ldots, n$ with bounded 4th moments, $s_i$ is a nonstochastic vector that weights the $v_i$'s, and $A$ is a conformable matrix.

**Lemma 5.** *Suppose:*

1. $\mathrm{Var}(T)^{-1/2}$ *is bounded;*
2. $\sum_i s_{il}^4 \to 0$*; and*
3. $||G_L G_L'||_F + ||G_U G_U'||_F \to 0$*, 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 $G_{U,ij} = G_{ij} 1\{i < j\}$.*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\delta_v$ with $||\delta_v||_2 \leq C$,

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

and the same argument can be applied to the $G_U$ term. Finally,

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

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

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

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

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

which suffices for the result. □

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

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

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

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

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

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

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

$$\begin{aligned}
\frac{1}{\sqrt{K}} \sum_i \sum_{j\neq i} G_{ij} e_i X_j &= \frac{1}{\sqrt{K}} \sum_i \sum_{j\neq i} G_{ij} \left(\nu_i R_j + \nu_i \eta_j + R_{\Delta i} R_j + R_{\Delta i}\eta_j\right), \text{ and} \\
\sum_i \sum_{j\neq i} G_{ij} R_{\Delta i} R_j &= \sum_i \sum_{j\neq i} G_{ij} \left(R_{Yi} - R_i \beta\right) R_j \\
&= \sum_i \sum_{j\neq i} G_{ij} R_{Yi} R_j - \sum_i \sum_{j\neq i} G_{ij} R_i R_j \left(\frac{\sum_i \sum_{j\neq i} G_{ij} R_{Yi} R_j}{\sum_i \sum_{j\neq i} G_{ij} R_i R_j}\right) = 0,
\end{aligned}$$

while $T_{XX}$ is immediate.

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

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

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

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

so that

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

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. $\frac{1}{r_n^2} \sum_i s_{il}^4 \to 0$ for all $l$; and
3. $||G_L G_L'||_F + ||G_U G_U'||_F \to 0$, where $G_L$ is a lower-triangular matrix with elements $G_{L,ij} = \frac{1}{\sqrt{r_n}} G_{ij} 1\{i > j\}$ and $G_U$ is an upper-triangular matrix with elements $G_{U,ij} = \frac{1}{\sqrt{r_n}} G_{ij} 1\{i < j\}$.

Condition (2) follows from Assumption 1(d) and applying the Cauchy-Schwarz inequality. Condition (3) is immediate from Assumption 1(e). For Condition (1), I show that Assumption 1(b) and (c) imply that, for any nonstochastic scalars $c_1, c_2, c_3$ that are finite and not all 0, $\text{Var}(T)^{-1/2}$ is bounded. Since $Cov\left(\sum_i s_i' v_i, \sum_i \sum_{j \neq i} G_{ij} v_i' A v_j\right) = 0$,

$$\text{Var}\,(T) = \frac{1}{r_n} \text{Var}\left(\sum_i s_i' v_i\right) + \frac{1}{r_n} \text{Var}\left(\sum_i \sum_{j \neq i} G_{ij} v_i' A v_j\right), \tag{21}$$

so it suffices to show that either term is bounded below. The second term is:

$$\begin{aligned} Var\left(\sum_i \sum_{j \neq i} G_{ij} v_i' A v_j\right) &= \sum_i \sum_{j \neq i} \sum_k \sum_{l \neq k} G_{ij} G_{kl} E\left[v_i' A v_j v_k' A v_l\right] \\ &= \sum_i \sum_{j \neq i} G_{ij}^2 E\left[v_i' A v_j v_j' A' v_i\right] + \sum_i \sum_{j \neq i} G_{ij} G_{ji} E\left[v_i' A v_j v_j' A v_i\right]. \end{aligned}$$

It can be shown that:

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

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

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

$$\begin{aligned} \text{Var}\left(\sum_i \sum_{j \neq i} G_{ij} v_i' A v_j\right) &= \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(c_3^2 E\left[\eta_i^2\right] E\left[\eta_j^2\right] + c_2^2 E\left[\nu_i^2\right] E\left[\eta_j^2\right] + c_1^2 E\left[\nu_i^2\right] E\left[\nu_j^2\right]\right) \\ &+ \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(2 c_1 c_3 E\left[\nu_i \eta_i\right] E\left[\nu_j \eta_j\right] + 2 c_1 c_2 E\left[\nu_i^2\right] E\left[\nu_j \eta_j\right] + 2 c_2 c_3 E\left[\eta_i \nu_i\right] E\left[\eta_j^2\right]\right) \\ &= \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(c_3^2\left(1 - \rho_i^2\right) E\left[\eta_i^2\right] E\left[\eta_j^2\right] + c_1^2\left(1 - \rho_j^2\right) E\left[\nu_i^2\right] E\left[\nu_j^2\right]\right) \\ &+ \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(c_3^2 \rho_i^2 E\left[\eta_i^2\right] E\left[\eta_j^2\right] + c_2^2 E\left[\nu_i^2\right] E\left[\eta_j^2\right] + c_1^2 \rho_j^2 E\left[\nu_i^2\right] E\left[\nu_j^2\right]\right) \\ &+ \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(2 c_1 c_3 \rho_i \sqrt{E\left[\eta_i^2\right] E\left[\nu_i^2\right]} \rho_j \sqrt{E\left[\eta_j^2\right] E\left[\nu_j^2\right]} + 2 c_1 c_2 E\left[\nu_i^2\right] \rho_j \sqrt{E\left[\eta_j^2\right] E\left[\nu_j^2\right]}\right) \\ &+ \sum_i \sum_{j \neq i} \left(G_{ij}^2 + G_{ij} G_{ji}\right)\left(2 c_2 c_3 \rho_i \sqrt{E\left[\eta_i^2\right] E\left[\nu_i^2\right]} E\left[\eta_j^2\right]\right) \end{aligned}$$

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

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 $\sum_i \sum_{j\neq i}\left(G_{ij}^2 + G_{ij}G_{ji}\right) \geq 0$. The inequality in the final line first applies Assumption 1(b). Using a similar argument,

$$
\begin{aligned}
\operatorname{Var}\left(\sum_i s_i' v_i\right) &= \sum_i s_i' \operatorname{Var}\left(v_i\right) s_i = \sum_i s_{i1}^2 E[\eta_i]^2 + 2 s_{i1}s_{i2}E[\eta_i\nu_i] + s_{i2}^2 E[\nu_i^2] \\
&= \sum_i (1-\rho_i)^2 E[\eta_i^2]s_{i1}^2 + \left(\rho_i s_{i1}\sqrt{E[\eta_i^2]} + s_{i2}\sqrt{E[\nu_i^2]}\right)^2 \geq \sum_i (1-\rho_i)^2 E[\eta_i^2] s_{i1}^2.
\end{aligned}
$$

A similar argument yields $\operatorname{Var}\left(\sum_i s_i' v_i\right) \geq \sum_i (1-\rho_i)^2 E[\eta_i^2] s_{i2}^2$. Due to Assumption 1(c), at least one of the following must hold: (i) $\frac{1}{r_n}\sum_i \sum_{j\neq i}\left(G_{ij}^2 + G_{ij}G_{ji}\right) \geq \underline{c}$ (ii) $\frac{1}{r_n}\sum_i s_{i1}^2 \geq \underline{c}$, or (iii) $\frac{1}{r_n}\sum_i s_{i2}^2 \geq \underline{c}$. Hence, $\operatorname{Var}(T)^{-1/2}$ is bounded.

Finally, since $\nu_i, \eta_i$ are mean zero, the expectations are immediate: $E\left[T_{ee}\right] = \sum_i \sum_{j\neq i} G_{ij}R_{\Delta j}R_{\Delta i}$ and $E\left[T_{XX}\right] = \sum_i \sum_{j\neq i} G_{ij}R_jR_i$. □

## B.2 Proofs for Section 3.2

*Proof of Equation* (8). Expanding the variance,

$$
\begin{aligned}
\operatorname{Var}\left(\sum_i \sum_{j\neq i} G_{ij}e_iX_j\right) &= E\left[\left(\sum_i \sum_{j\neq i} G_{ij}e_iX_j\right)^2\right] = E\left[\sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}e_iX_jG_{kl}e_kX_l\right] \\
&= \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[\nu_iX_j\nu_kX_l\right] + \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[\nu_iX_jR_{\Delta k}X_l\right] \\
&\quad + \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[R_{\Delta i}X_j\nu_kX_l\right] + \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[R_{\Delta i}X_jR_{\Delta k}X_l\right]
\end{aligned}
$$

The first term is:

$$
\begin{aligned}
&\sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[\nu_iX_j\nu_kX_l\right] \\
&\quad= \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[\nu_iR_j\nu_kR_l + \nu_i\eta_j\nu_kR_l + \nu_iR_j\nu_k\eta_l + \nu_i\eta_j\nu_k\eta_l\right] \\
&\quad= \sum_i \sum_{j\neq i}\sum_k \sum_{l\neq k} G_{ij}G_{kl}E\left[\nu_iR_j\nu_kR_l + \nu_i\eta_j\nu_k\eta_l\right] \\
&\quad= \sum_i \sum_{j\neq i}\sum_{k\neq i} E\left[\nu_i^2\right]G_{ij}G_{ik}R_jR_k + \sum_i \sum_{j\neq i}\left(\sum_{l\neq i} G_{ij}G_{il}E\left[\nu_i\eta_j\nu_i\eta_l\right] + \sum_{l\neq j} G_{ij}G_{jl}E\left[\nu_i\eta_j\nu_j\eta_l\right]\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_i \sum_{j\neq i}\left(\sum_{k\neq i,j} G_{ij}G_{ki}E\left[\nu_i\eta_j\nu_k\eta_i\right]+\sum_{k\neq i,j} G_{ij}G_{kj}E\left[\nu_i\eta_j\nu_k\eta_j\right]\right)\\
&=\sum_i \sum_{j\neq i}\sum_{k\neq i} E\left[\nu_i^2\right] G_{ij}G_{ik}R_jR_k+\sum_i\sum_{j\neq i}\left(G_{ij}^2E\left[\nu_i^2\eta_j^2\right]+G_{ij}G_{ji}E\left[\nu_i\eta_i\eta_j\nu_j\right]\right)\\
&=\sum_i \sum_{j\neq i}\sum_{k\neq i} E\left[\nu_i^2\right] G_{ij}G_{ik}R_jR_k+\sum_i\sum_{j\neq i}\left(G_{ij}^2E\left[\nu_i^2\right]E\left[\eta_j^2\right]+G_{ij}G_{ji}E\left[\nu_i\eta_i\right]E\left[\eta_j\nu_j\right]\right)
\end{aligned}
$$

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:

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

The expression stated in the equation combines these expressions. □

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 $h(.)$ form so that Assumption 3 inequalities can be applied. These inequalities are also sufficiently general that other components of the variance matrix in (7) 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. Following Equation (6), $r_n:=\sum_i\tilde{R}_i^2+\sum_i\tilde{R}_{\Delta i}^2+\sum_i\sum_{j\neq i}G_{ij}^2$. Let $C_i, C_{ij}, C_{ijk}$ denote nonstochastic objects that are non-negative and are bounded above by $C$. I use $h_4^A(.)$ and $h_4^B(.)$ to denote two different functions that satisfy the above definition for $h_4$.

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

*(a)* $\left|\sum_{i\neq j\neq k}^n C_{ijk}\left(\sum_{l\neq i,j,k}h_4^A(i,j,k,l)R_{ml}\right)\left(\sum_{l\neq i,j,k}h_4^B(i,j,k,l)R_{ml}\right)\right|\leq C\sum_i\tilde{R}_{mi}^2,$
$\left|\sum_{i\neq j}^n C_{ij}\left(\sum_{k\neq i,j}\sum_{l\neq i,j,k}h_4^A(i,j,k,l)R_{ml}\right)\left(\sum_{k\neq i,j}\sum_{l\neq i,j,k}h_4^B(i,j,k,l)R_{ml}\right)\right|\leq C\sum_i\tilde{R}_{mi}^2,$
*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)* $\left|\sum_{i\neq j}^n C_{ij}\left(\sum_{k\neq i,j}h_3^A(i,j,k)R_{mk}\right)\left(\sum_{k\neq i,j}h_3^B(i,j,k)R_{mk}\right)\right|\leq C\sum_i\tilde{R}_{mi}^2$
*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 (c). By applying the Cauchy-Schwarz inequality,

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

$$\leq \max_i C_i \left( \sum_i \left( \sum_{j \neq i} h_2^A (i,j) R_{mj} \right)^2 \right)^{1/2} \left( \sum_i \left( \sum_{j \neq i} h_2^B (i,j) R_{mj} \right)^2 \right)^{1/2}$$

$$\leq \max_i C_i \left( \sum_i \tilde{R}_{mi}^2 \right)^{1/2} \left( \sum_i \tilde{R}_{mi}^2 \right)^{1/2} \leq C \sum_i \tilde{R}_{mi}^2.$$

The proof of all other parts are entirely analogous. ☐

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

*(a)* $\text{Var}\left( \sum_{i \neq j}^n G_{ij} F_{ij} V_{1i} V_{2i} V_{3j} V_{4j} \right) \leq C r_n$.

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

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

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

*Proof of Lemma 7.* **Proof of Lemma 7(a).**
Using the decomposition from AS23,

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

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

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

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

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

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,

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

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:

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

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

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

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

To deal with 3 stochastic terms,

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

$$\leq \sum_{i\neq j}^{n} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\left(\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)\left[\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k} + \sum_{k\neq i,j} G_{ki}F_{ki}\check{M}_{kj,-ki}R_{1k}\right]$$

$$+ \max_{i,j} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\sum_{i\neq j}^{n}\sum_{l\neq i,j}\left(\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)\left(\sum_{k\neq i,l} G_{kl}F_{kl}\check{M}_{ki,-kl}R_{1k}\right)$$

$$\leq \sum_{i}\sum_{j\neq i} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\left(\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)\left[\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k} + \sum_{k\neq i,j} G_{ki}F_{ki}\check{M}_{kj,-ki}R_{1k}\right]$$

$$+ \max_{i,j} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\sum_{i}\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)^{2} - \max_{i,j} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\sum_{i}\sum_{j\neq i}\left(\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)^{2}$$

$$\leq \max_{i,j} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\sum_{i}\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)^{2}$$

$$+ \sum_{i\neq j}^{n} \operatorname{Var}\left(v_{2i}V_{3j}V_{4j}\right)\left(\sum_{k\neq i,j} G_{kj}F_{kj}\check{M}_{ki,-kj}R_{1k}\right)\left(\sum_{k\neq i,j} G_{ki}F_{ki}\check{M}_{kj,-ki}R_{1k}\right)$$

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,

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

$$\leq \max_{i,j} \operatorname{Var}\left(v_{1i}V_{3j}V_{4j}\right)\sum_{i}\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}F_{ij}\check{M}_{ik,-ij}R_{2k}\right)^{2}$$

$$+ \sum_{i}\sum_{j\neq i} \operatorname{Var}\left(v_{1i}V_{3j}V_{4j}\right)\left(\sum_{k\neq i,j} G_{ij}F_{ij}\check{M}_{ik,-ij}R_{2k}\right)\left(\sum_{k\neq i,j} G_{ji}F_{ji}\check{M}_{jk,-ij}R_{2k}\right).$$

since the expansion in the intermediate steps are entirely analogous.

Turning to the sum with two stochastic objects,

$$\operatorname{Var}\left(\sum_{i\neq j\neq k}^{n} G_{ij}F_{ij}\check{M}_{ik,-ij}R_{1i}R_{2k}V_{3j}V_{4j}\right) = \operatorname{Var}\left(\sum_{i} V_{3i}V_{4i}\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{ji}F_{ji}\check{M}_{jk,-ij}R_{1j}R_{2k}\right)\right)$$

$$= \sum_{i} \operatorname{Var}\left(V_{3i}V_{4i}\right)\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{ji}F_{ji}\check{M}_{jk,-ij}R_{1j}R_{2k}\right)^{2} \leq \max_{i} \operatorname{Var}\left(V_{3i}V_{4i}\right)\sum_{i}\left(\sum_{j\neq i}\sum_{k\neq i,j} G_{ji}F_{ji}\check{M}_{jk,-ij}R_{1j}R_{2k}\right)^{2}.$$

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

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

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

With four stochastic objects,

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i\neq j\neq l}^{n} G_{ij}F_{ij}\check{M}_{jl,-ij}V_{1i}V_{2i}v_{3j}v_{4l}\right)\\
&\leq \max_{i,j,k}\operatorname{Var}\left(V_{1i}V_{2i}v_{3j}v_{4l}\right)\sum_{i\neq j\neq l}^{n}\sum_{i_2\neq i,j,l}\left(G_{ij}F_{ij}\check{M}_{jl,-ij}G_{i_2j}F_{i_2j}\check{M}_{jl,-i_2j} + G_{ij}F_{ij}\check{M}_{jl,-ij}G_{i_2l}F_{i_2l}\check{M}_{lj,-i_2l}\right)\\
&\quad+\max_{i,j,k}\operatorname{Var}\left(V_{1i}V_{2i}v_{3j}v_{4l}\right)3!\sum_{i\neq j\neq l}^{n} G_{ij}^2F_{ij}^2\check{M}_{jl,-ij}^2.
\end{aligned}$$

Simplifying the first line,

$$\begin{aligned}
&\sum_{i\neq j\neq l}^{n}\sum_{i_2\neq i,j,l}\left(G_{ij}F_{ij}\check{M}_{jl,-ij}G_{i_2j}F_{i_2j}\check{M}_{jl,-i_2j} + G_{ij}F_{ij}\check{M}_{jl,-ij}G_{i_2l}F_{i_2l}\check{M}_{lj,-i_2l}\right)\\
&\leq\left(\sum_{i\neq j\neq l\neq i_2}^{n} G_{ij}^2G_{i_2j}^2\check{M}_{jl,-ij}^2\right)^{1/2}\left(\sum_{i\neq j\neq l\neq i_2}^{n} F_{ij}^2F_{i_2j}^2\check{M}_{jl,-i_2j}^2\right)^{1/2}\\
&\quad+\left(\sum_{i\neq j\neq l\neq i_2}^{n} G_{ij}^2G_{i_2l}^2\check{M}_{jl,-ij}^2\right)^{1/2}\left(\sum_{i\neq j\neq l\neq i_2}^{n} F_{ij}^2F_{i_2l}^2\check{M}_{lj,-i_2j}^2\right)^{1/2}.
\end{aligned}$$

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

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

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

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

These inequalities suffice for the result due to Assumption 3.

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

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

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,

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

By using the same expansion step as before,

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

The $\sum_{i\neq j\neq k\neq i_2}^{n}\left(G_{ij}F_{ij}G_{i_2j}F_{i_2j}\check{M}_{ik,-ij}\check{M}_{i_2k,-ij}+G_{ij}F_{ij}G_{i_2k}F_{i_2k}\check{M}_{ij,-ik}\check{M}_{i_2j,-ik}\right)$ term has the required order due to the same argument as the proof of Lemma 7(b). Next,

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{i\neq j\neq k\neq l}^{n} G_{ij}F_{ij}\check{M}_{ik,-ij}\check{M}_{jl,-ijk}V_{1i}v_{2k}R_{3j}v_{4l}\right) = \operatorname{Var}\left(\sum_{i\neq j\neq k}^{n} V_{1i}v_{2k}v_{4j}\left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)\right) \\
&\leq \max_{i,j,k}\operatorname{Var}\left(V_{1i}v_{2k}v_{4j}\right)\sum_{i\neq j\neq k}^{n}\sum_{i_2\neq i,j,k}\left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)\left(\sum_{l\neq i_2,j,k} G_{i_2l}F_{i_2l}\check{M}_{i_2k,-i_2l}\check{M}_{lj,-i_2lk}R_{3l}\right)
\end{aligned}$$

$$+\max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) \sum_{i\neq j\neq k}^{n} \sum_{i_2\neq i,j,k} \left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)\left(\sum_{l\neq i_2,j,k} G_{i_2l}F_{i_2l}\check{M}_{i_2j,-i_2l}\check{M}_{lk,-i_2lj}R_{3l}\right)$$

$$+\max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) 3! \sum_{i\neq j\neq k}^{n} \check{M}_{ik,-ij}^2 \left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)^2$$

$$\leq \max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) \sum_{k}\sum_{j\neq k} \left(\sum_{i\neq k,j}\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)^2$$

$$-\max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) \sum_{k}\sum_{j\neq k}\sum_{i\neq k,j} \left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)^2$$

$$+\max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) \sum_{k}\sum_{j\neq k} \left(\sum_{i\neq k,j}\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)\left(\sum_{i\neq k,j}\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ij,-il}\check{M}_{lk,-ilj}R_{3l}\right)$$

$$-\max_{i,j,k} \mathrm{Var}\left(V_{1i}v_{2k}v_{4j}\right) \sum_{k}\sum_{j\neq k}\sum_{i\neq k,j} \left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ik,-il}\check{M}_{lj,-ilk}R_{3l}\right)\left(\sum_{l\neq i,j,k} G_{il}F_{il}\check{M}_{ij,-il}\check{M}_{lk,-ilj}R_{3l}\right)$$

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

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

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

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

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

$$+\max_{i,j} \mathrm{Var}\left(V_{1i}v_{2j}\right)\sum_{i\neq j\neq i_2}^{n}\left(G_{ij}F_{ij}\sum_{k\neq i,j}\sum_{l\neq i,j,k}\check{M}_{ij,-ik}\check{M}_{kl,-ijk}R_{3k}R_{4l}\right)$$

$$\left(G_{i_2j}F_{i_2j}\sum_{k\neq i_2,j}\sum_{l\neq i_2,j,k}\check{M}_{i_2j,-i_2k}\check{M}_{kl,-i_2jk}R_{3k}R_{4l}\right)$$

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

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

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

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

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

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

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

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

$$\le \max_{i,j,k} \mathrm{Var}\left(V_{1i} v_{3j} v_{4k}\right) \left( \sum_{k \neq j}^{n} \left( \sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ij} F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2 - \sum_{k \neq j}^{n} \sum_{i \neq k,j} \left( \sum_{l \neq i,j,k} G_{ij} F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2 \right)$$

$$+ \max_{i,j,k} \mathrm{Var}\left(V_{1i} v_{3j} v_{4k}\right) \sum_{k} \sum_{j \neq k} \left( \sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ij} F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right) \left( \sum_{i \neq k,j} \sum_{l \neq i,j,k} G_{ik} F_{ik} \check{M}_{il,-ik} \check{M}_{kj,-ikl} R_{2l} \right)$$

$$- \max_{i,j,k} \mathrm{Var}\left(V_{1i} v_{3j} v_{4k}\right) \sum_{k} \sum_{j \neq k} \sum_{i \neq k,j} \left( \sum_{l \neq i,j,k} G_{ij} F_{ij} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right) \left( \sum_{l \neq i,j,k} G_{ik} F_{ik} \check{M}_{il,-ik} \check{M}_{kj,-ikl} R_{2l} \right)$$

$$+ \max_{i,j,k} \mathrm{Var}\left(V_{1i} v_{3j} v_{4k}\right) 3! \sum_{i \neq j \neq k}^{n} \left( G_{ij} F_{ij} \sum_{l \neq i,j,k} \check{M}_{il,-ij} \check{M}_{jk,-ijl} R_{2l} \right)^2 .$$

Next,

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

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

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

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

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

Finally,

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

□

**Lemma 8.** *Under Assumption 3, the following hold:*
*(a)* $\mathrm{Var}\left(\sum_{i\neq j\neq k}^{n} G_{ij}F_{ik}V_{1j}V_{2k}V_{3i}V_{4i}\right) \leq Cr_n$.

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

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

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

With four stochastic objects,

$$\begin{aligned}
Var\left(\sum_{i\neq j\neq k}^{n} G_{ij}F_{ik}V_{3i}V_{4i}v_{1j}v_{2k}\right) &\leq \max_{i,j,k}\mathrm{Var}\left(V_{3i}V_{4i}v_{1j}v_{2k}\right)\sum_{i\neq j\neq k}^{n}\sum_{i_2\neq i,j,k}\left(G_{ij}F_{ik}G_{i_2j}F_{i_2k} + G_{ij}F_{ik}G_{i_2k}F_{i_2j}\right) \\
&\quad + \max_{i,j,k}\mathrm{Var}\left(V_{1i}V_{2i}v_{3j}v_{4l}\right)3!\sum_{i\neq j\neq k}^{n} G_{ij}^2F_{ik}^2.
\end{aligned}$$

Observe that, due to Assumption 3(a),

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

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

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{i\neq j\neq k}^{n} G_{ij}F_{ik}V_{3i}V_{4i}R_{1j}v_{2k}\right) \\
&= \mathrm{Var}\left(\sum_{i}\sum_{j\neq i} F_{ij}V_{3i}V_{4i}v_{2j}\left(\sum_{k\neq i,j} G_{ik}R_{1k}\right)\right) \\
&\leq \sum_{i}\sum_{j\neq i}\mathrm{Var}\left(V_{3i}V_{4i}v_{2j}\right)\left(\sum_{k\neq i,j} F_{ij}G_{ik}R_{1k}\right)\left[\sum_{k\neq i,j} F_{ij}G_{ik}R_{1k} + \sum_{k\neq i,j} F_{ji}G_{jk}R_{1k}\right] \\
&+ \max_{i,j}\mathrm{Var}\left(V_{3i}V_{4i}v_{2j}\right)\sum_{i}\sum_{j\neq i}\sum_{i_2\neq i,j}\left(\sum_{k\neq i,j} F_{ij}G_{ik}R_{1k}\right)\left(\sum_{k\neq i_2,l} F_{i_2j}G_{i_2k}R_{1k}\right) \\
&\leq \max_{i,j}\mathrm{Var}\left(V_{3i}V_{4i}v_{2j}\right)\sum_{i}\left(\sum_{j\neq i}\sum_{k\neq i,j} F_{ij}G_{ik}R_{1k}\right)^2 + \sum_{i}\sum_{j\neq i}\mathrm{Var}\left(V_{3i}V_{4i}v_{2j}\right)\left(\sum_{k\neq i,j} F_{ij}G_{ik}R_{1k}\right)\left(\sum_{k\neq i,j} F_{ji}G_{jk}R_{1k}\right).
\end{aligned}$$

Similarly,

$$\mathrm{Var}\left(\sum_{i\neq j\neq k}^{n} G_{ij}F_{ik}V_{3i}V_{4i}v_{1j}R_{2k}\right) = \mathrm{Var}\left(\sum_{i}\sum_{j\neq i} V_{3i}V_{4i}v_{1j}\left(\sum_{k\neq i,j} G_{ij}F_{ik}R_{2k}\right)\right)$$

$$\leq \max_{i,j} \operatorname{Var}\left(V_{3i}V_{4i}v_{1j}\right) \sum_{i} \left( \left( \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}F_{ik}R_{2k} \right)^2 + \sum_{j\neq i} \left( \sum_{k\neq i,j} G_{ij}F_{ik}R_{2k} \right) \left( \sum_{k\neq i,j} G_{ji}F_{jk}R_{2k} \right) \right).$$

Turning to the sum with two stochastic objects,

$$\operatorname{Var}\left( \sum_{i\neq j\neq k}^{n} G_{ij}F_{ik}V_{3i}V_{4i}R_{1j}R_{2k} \right) = \operatorname{Var}\left( \sum_{i} V_{3i}V_{4i} \left( \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}F_{ik}R_{1j}R_{2k} \right) \right)$$

$$\leq \max_{i} \operatorname{Var}\left(V_{3i}V_{4i}\right) \sum_{i} \left( \sum_{j\neq i} \sum_{k\neq i,j} G_{ij}F_{ik}R_{1j}R_{2k} \right)^2 .$$

**Proof of Lemma 8(b).**
Decompose the term:

$$\sum_{i\neq j\neq k\neq l}^{n} G_{ij}F_{ik}\check{M}_{il,-ijk}V_{1j}V_{2k}V_{3i}V_{4l}$$

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

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

Consider the $v_{1j}$ line first.

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

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

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

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

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

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

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

and

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

$$\leq \max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{1j}v_{4k}\right)\sum_{k}\sum_{j\neq k}\left(\left(\sum_{i\neq j,k}\sum_{l\neq i,j,k}G_{ij}F_{ik}\check{M}_{ik,-ijl}R_{2l}\right)^{2}-\sum_{i\neq j,k}\left(\sum_{l\neq i,j,k}G_{ij}F_{ik}\check{M}_{ik,-ijl}R_{2l}\right)^{2}\right)$$

$$+\max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{1j}v_{2k}\right)\sum_{k}\sum_{j\neq k}\left(\sum_{i\neq j,k}\sum_{l\neq i,j,k}G_{ij}F_{ik}\check{M}_{ik,-ijl}R_{2l}\right)\left(\sum_{i\neq j,k}\sum_{l\neq i,j,k}G_{ik}F_{ij}\check{M}_{il,-ijk}R_{2l}\right)$$

$$-\max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{1j}v_{2k}\right)\sum_{k}\sum_{j\neq k}\sum_{i\neq j,k}\left(\sum_{l\neq i,j,k}G_{ij}F_{ik}\check{M}_{ik,-ijl}R_{2l}\right)\left(\sum_{l\neq i,j,k}G_{ik}F_{ij}\check{M}_{il,-ijk}R_{2l}\right)$$

$$+\max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{1j}v_{4k}\right)3!\sum_{i\neq j\neq k}^{n}G_{ij}^{2}F_{ik}^{2}\left(\sum_{l\neq i,j,k}\check{M}_{ik,-ijl}R_{2l}\right)^{2}$$

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

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

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

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

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

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

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

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

$$+\max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{2k}v_{4j}\right)\sum_{j}\sum_{k\neq j}\left(\sum_{i\neq j,k}\sum_{l\neq i,j,k}G_{il}F_{ik}\check{M}_{ij,-ilk}R_{1l}\right)\left(\sum_{i\neq j,k}\sum_{l\neq i,j,k}G_{il}F_{ij}\check{M}_{ik,-ilj}R_{1l}\right)$$

$$-\max_{i,j,k}\mathrm{Var}\left(V_{3i}v_{2k}v_{4j}\right)\sum_{j}\sum_{k\neq j}\sum_{i\neq j,k}\left(\sum_{l\neq i,j,k}G_{il}F_{ik}\check{M}_{ij,-ilk}R_{1l}\right)\left(\sum_{l\neq i,j,k}G_{il}F_{ij}\check{M}_{ik,-ilj}R_{1l}\right)$$

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

and

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

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

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

□

**Lemma 9.** *Under Assumption 3, the following hold:*

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

(b) $\operatorname{Var}\left(\sum_{i\neq j\neq k}^{n}G_{ji}^{2}\check{M}_{ik,-ij}V_{1i}V_{2k}V_{3j}V_{4j}\right)\leq Cr_n$;

(c) $\operatorname{Var}\left(\sum_{i\neq j\neq l}^{n}G_{ji}^{2}\check{M}_{jl,-ij}V_{1i}V_{2i}V_{3j}V_{4l}\right)\leq Cr_n$;

(d) $\operatorname{Var}\left(\sum_{i\neq j\neq k\neq l}^{n}G_{ji}^{2}V_{1i}\check{M}_{ik,-ij}V_{2k}V_{3j}\check{M}_{jl,-ijk}V_{4l}\right)\leq Cr_n$;

(e) $\operatorname{Var}\left(\sum_{i\neq j\neq k}^{n}G_{ji}F_{ki}V_{1j}V_{2k}V_{3i}V_{4i}\right)\leq Cr_n$;

(f) $\operatorname{Var}\left(\sum_{i\neq j\neq k\neq l}^{n}G_{ji}F_{ki}\check{M}_{il,-ijk}V_{1j}V_{2k}V_{3i}V_{4l}\right)\leq Cr_n$.

*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 $G_{ij}$. □

*Proof of Theorem 2.* **Proof of Unbiasedness**

The variance expression can be equivalently be written as:

$$V_{LM}=\sum_{i}\left(E\left[\nu_i^2\right]\left(\sum_{j\neq i}G_{ij}R_j\right)^{2}+2\left(\sum_{j\neq i}G_{ij}R_j\right)\left(\sum_{j\neq i}G_{ji}R_{\Delta j}\right)E\left[\nu_i\eta_i\right]+E\left[\eta_i^2\right]\left(\sum_{j\neq i}G_{ji}R_{\Delta j}\right)^{2}\right)$$
$$+\sum_{i}\sum_{j\neq i}G_{ij}^{2}E\left[\nu_i^2\right]E\left[\eta_j^2\right]+\sum_{i}\sum_{j\neq i}G_{ij}G_{ji}E\left[\eta_i\nu_i\right]E\left[\eta_j\nu_j\right].\tag{22}$$

To ease notation, let:

$$
\begin{aligned}
A_{1i} &:= \sum_{j\neq i}\sum_{k\neq i} G_{ij}X_jG_{ik}X_ke_i\left(e_i - Q_i'\hat{\tau}_{\Delta,-ijk}\right),\\
A_{2i} &:= \sum_{j\neq i}\sum_{k\neq i} G_{ij}X_jG_{ki}e_ke_i\left(X_i - Q_i'\hat{\tau}_{-ijk}\right),\\
A_{3i} &:= \sum_{j\neq i}\sum_{k\neq i} G_{ji}e_jG_{ki}e_kX_i\left(X_i - Q_i'\hat{\tau}_{-ijk}\right),\\
A_{4ij} &:= X_i\sum_{k\neq j} \check{M}_{ik,-ij}X_ke_j\left(e_j - Q_j'\hat{\tau}_{\Delta,-ijk}\right), \text{ and}\\
A_{5ij} &:= e_i\sum_{k\neq j} \check{M}_{ik,-ij}X_ke_j\left(X_j - Q_j'\hat{\tau}_{-ijk}\right).
\end{aligned}
$$

Take expectation of $A_1$:

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

Evaluating the first term,

$$
\begin{aligned}
&\sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}E\left[X_j\right]G_{ik}E\left[X_k\right]E\left[e_i\left(e_i - Q_i'\hat{\tau}_{\Delta,-ijk}\right)\right]\\
&= \sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}R_jG_{ik}R_k\left(E\left[e_i^2\right] - E\left[e_iQ_i'\hat{\tau}_{\Delta,-ijk}\right]\right) = \sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}R_jG_{ik}R_k\left(E\left[e_i^2\right] - E\left[e_iQ_i'\tau_{\Delta}\right]\right)\\
&= \sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}R_jG_{ik}R_kE\left[e_i\nu_i\right] = \sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}R_jG_{ik}R_kE\left[\nu_i^2\right].
\end{aligned}
$$

Using an analogous argument for the second term,

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

Combining them,

$$
\begin{aligned}
&E\left[\sum_i\sum_{j\neq i}\sum_{k\neq i} G_{ij}X_jG_{ik}X_ke_i\left(e_i - Q_i'\hat{\tau}_{\Delta,-ijk}\right)\right] = \sum_i\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}R_jG_{ik}R_kE\left[\nu_i^2\right] + \sum_i\sum_{j\neq i} G_{ij}^2\left(R_j^2 + E\left[\eta_j^2\right]\right)E\left[\nu_i^2\right]\\
&= \sum_i\sum_{j\neq i}\sum_{k\neq i} G_{ij}R_jG_{ik}R_kE\left[\nu_i^2\right] + \sum_i\sum_{j\neq i} G_{ij}^2E\left[\nu_i^2\right]E\left[\eta_j^2\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E\left[A_{2i}\right] &= \left(\sum_{j\neq i} G_{ij}R_j\right)\left(\sum_{j\neq i} G_{ji}R_{\Delta j}\right)E\left[\nu_i\eta_i\right] + \sum_{j\neq i} G_{ij}G_{ji}E\left[\eta_i\nu_i\right]E\left[\eta_j\nu_j\right], \text{ and}\\
E\left[A_{3i}\right] &= E\left[\eta_i^2\right]\left(\sum_{j\neq i} G_{ji}R_{\Delta j}\right)^2 + \sum_{j\neq i} G_{ji}^2E\left[\eta_i^2\right]E\left[\nu_j^2\right].
\end{aligned}
$$

For the $A_4$ and $A_5$ terms, observe that:

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

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

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

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

**Proof of Consistency**

By Chebyshev's inequality,

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

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 Appendix A.1, let $\tilde{R}_i := \sum_{j\neq i} G_{ij} R_j$ and $\tilde{R}_{\Delta i} := \sum_{j\neq i} G_{ji} R_{\Delta j}$. Further, to simplify notation, let $\rho_i := corr(\eta_i \nu_i)$.

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 (22) 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:

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

and the second line is:

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

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

Consequently,

$$
V_{LM} \succeq \sum_i \tilde{R}_i^2 + \sum_i \tilde{R}_{\Delta i}^2 + \sum_i\sum_{j\neq i} G_{ij}^2 =: r_n. \tag{23}
$$

Hence, since $r_n \to \infty$ by Assumption 1(d), $V_{LM}$ also diverges. 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 (23)). If this is true, then since the denominator has order at least $r_n^2$, the variance estimator is consistent. The A1 and A2 terms have the form:

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

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:

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

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. □

# C Proofs for Section 4

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

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

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

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

Then, the conditional normal distribution is:

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

Hence, $PX$ and $PY$ (i.e, $Z'X$, $Z'Y$) are sufficient statistics for $\pi_Y, \pi$.

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, $s_1^{*\prime}s_1^* = s_1'F'Fs_1 = s_1's_1$. Invariance of $(s_1's_2, s_2's_2)$ is analogous. Maximality states that if $s_1^{*\prime}s_1^* = s_1's_1$, 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 $s_1^{*\prime}s_1^* = s_1's_1$. Since $G$ is not an orthogonal matrix, $G'G \neq I$. Hence, $s_1^{*\prime}s_1^* = s_1'G'Gs_1 \neq s_1's_1$, a contradiction. To obtain the distribution,

$$\left[\begin{array}{c} s_1 \\ s_2 \end{array}\right] = \left[\begin{array}{c} \left(Z'Z\right)^{-1/2} Z'\left(Z\pi_Y + \zeta\right) \\ \left(Z'Z\right)^{-1/2} Z'\left(Z\pi + \eta\right) \end{array}\right] = \left[\begin{array}{c} \left(Z'Z\right)^{1/2}\pi_Y \\ \left(Z'Z\right)^{1/2}\pi \end{array}\right] + \left[\begin{array}{c} \left(Z'Z\right)^{-1/2} Z'\zeta \\ \left(Z'Z\right)^{-1/2} Z'\eta \end{array}\right].$$

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

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

$\square$

*Proof of Proposition 1.* Let

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

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

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

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

$$\left(\begin{array}{c} z_{1k} \\ z_{2k} \end{array}\right) \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left[\begin{array}{cc} \omega_{\zeta\zeta} & \omega_{\zeta\eta} \\ \omega_{\zeta\eta} & \omega_{\eta\eta} \end{array}\right]\right), \text{ and}$$

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

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

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

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

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

$$A = \begin{pmatrix} 0 & 2 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 1 & 0 \end{pmatrix}.$$

This means:

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

Let $\Sigma = AVA'$ to obtain the result as stated. To derive $\Sigma$ explicitly, I derive $V$ by applying the Isserlis' Theroem. 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 corrolary is that if $n$ is odd, then there is no such pairing, so the moment is always zero. Hence,

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

Similarly,

$$\text{Var}\left(z_{2k}^2\right) = E\left[z_{2k}^4\right] - \omega_{\eta\eta}^2 = 3\omega_{\eta\eta}^2 - \omega_{\eta\eta}^2 = 2\omega_{\eta\eta}^2,$$
$$Cov(z_{1k}, z_{1k}z_{2k}) = E\left[z_{1k}^2 z_{2k}\right] - E[z_{1k}]E[z_{1k}z_{2k}] = 0,$$

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

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

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

*Proof of Proposition 2.* Fix any alternative $\left(\pi^A, \pi_Y^A\right) \in \mathcal{S}$ with a corresponding $\left(\mu_1^A, \mu_2^A, \mu_3^A\right)$. Due to the restriction in $\mathcal{S}$,

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

is in the null space. The Neyman-Pearson test for $\mu^H$ against $\mu^A$ rejects large values of:

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

Hence, the most powerful test rejects large values of $X_2$, which is what LM does. By Lehmann and Romano (2005) Theroem 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 $\left(\pi^A, \pi_Y^A\right)$. □

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

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

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, consider the following transformation $L$:

$$L := \begin{pmatrix} \sqrt{\frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}} & -\frac{\sigma_{12}}{\sigma_{22}}\sqrt{\frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}} & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & 0 \\ 0 & -\frac{\sigma_{23}}{\sigma_{22}}\sqrt{\frac{\sigma_{22}}{\sigma_{33}\sigma_{22}-\sigma_{23}^2}} & \sqrt{\frac{\sigma_{22}}{\sigma_{33}\sigma_{22}-\sigma_{23}^2}} \end{pmatrix}.$$

Then,

$$Lm \sim N\left(L\mu, \begin{pmatrix} 1 & 0 & \frac{\sigma_{13}\sigma_{22}-\sigma_{12}\sigma_{23}}{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)(\sigma_{33}\sigma_{22}-\sigma_{23}^2)} \\ 0 & 1 & 0 \\ \frac{\sigma_{13}\sigma_{22}-\sigma_{12}\sigma_{23}}{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)(\sigma_{33}\sigma_{22}-\sigma_{23}^2)} & 0 & 1 \end{pmatrix}\right).$$

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), by using $U = m_2$ and $T = \left(\sqrt{\frac{\sigma_{22}}{\sigma_{33}\sigma_{22}-\sigma_{23}^2}}m_3 - \frac{\sigma_{23}}{\sqrt{\sigma_{22}(\sigma_{33}\sigma_{22}-\sigma_{23}^2)}}m_2, \sqrt{\frac{\sigma_{22}}{\sigma_{11}\sigma_{22}-\sigma_{12}^2}}m_1 - \frac{\sigma_{12}}{\sqrt{\sigma_{22}(\sigma_{11}\sigma_{22}-\sigma_{12}^2)}}m_2\right)'$ in their notation. 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 $\pm 1.96$ 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. $\square$

# D Proofs for Section 5

*Proof of Lemma 2.* The $A$ expressions can be written as:

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

Since these terms have a quadratic form, the variance estimator is also quadratic in $\beta_0$, i.e.,

$$\hat{V}_{LM} = B_0 + B_1\beta_0 + B_2\beta_0^2,$$

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

$$B_0 = \sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq k} \check{M}_{il,-ijk} G_{ij} X_j G_{ik} X_k Y_i Y_l + 2\sum_i \sum_{j\neq i} \sum_{k\neq i} \sum_{l\neq k} \check{M}_{il,-ijk} G_{ij} X_j G_{ki} X_l Y_i Y_k$$

$$
+\sum_{i}\sum_{j\neq i}\sum_{k\neq i}\sum_{l\neq k}\check{M}_{il,-ijk}X_lG_{ji}G_{ki}X_iY_jY_k
$$

$$
-\sum_{i}\sum_{j\neq i}\sum_{k\neq j}\sum_{l\neq i,k}\check{M}_{jl,-ijk}\check{M}_{ik,-ij}G_{ji}^2X_iX_kY_jY_l-\sum_{i}\sum_{j\neq i}\sum_{k\neq j}\sum_{l\neq i,k}\check{M}_{ik,-ij}\check{M}_{jl,-ijk}G_{ij}G_{ji}X_kX_lY_iY_j
$$

$B_1$ and $B_2$ are analogous by collecting the coefficients on $\beta_0, \beta_0^2$ from expressions $A_1$ to $A_5$. The test does not reject:

$$
\frac{\left(KT_{YX}-KT_{XX}\beta_0\right)^2}{B_0+B_1\beta_0+B_2\beta_0^2}\leq q\Leftrightarrow\left(KT_{XX}^2-qB_2\right)\beta_0^2-\left(2KT_{YX}T_{XX}+qB_1\right)\beta_0+\left(KT_{YX}^2-qB_0\right)\leq 0.
$$

Solutions exist when:

$$
D:=\left(2KT_{YX}T_{XX}+qB_1\right)^2-4\left(KT_{XX}^2-qB_2\right)\left(KT_{YX}^2-qB_0\right)\geq 0.
$$

The rest of the lemma is immediate from properties of solving quadratic inequalities. □

# E Proofs for Appendix A

## E.1 Proofs for Appendix A.1

*Proof of Equation (16).*

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

□

*Proof of Proposition 4.* Use $n_q^Q$ and $n_w^W$ to denote the number of observations in the instrument and covariate groups respectively, so

$$\begin{aligned}
\mu_3 &= \sum_i \sum_{j \neq i} G_{ij} R_i R_j = \sum_q \frac{n_q^Q}{n_q^Q - 1} \sum_{i \in \mathcal{N}_q^Q} \sum_{j \in \mathcal{N}_q^Q, j \neq i} \frac{1}{n_q^Q} R_i R_j - \sum_w \frac{n_w^W}{n_w^W - 1} \sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} \frac{1}{n_w^W} R_i R_j \\
&= \sum_q \frac{1}{n_q^Q - 1} \sum_{i \in \mathcal{N}_q^Q} \sum_{j \in \mathcal{N}_q^Q, j \neq i} R_i R_j - \sum_w \frac{1}{n_w^W - 1} \sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} R_i R_j \\
&= \sum_w \left( \sum_{q \in \mathcal{M}_w} \frac{1}{n_q^Q - 1} \sum_{i \in \mathcal{N}_q^Q} \sum_{j \in \mathcal{N}_q^Q, j \neq i} R_i R_j - \frac{1}{n_w^W - 1} \sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} R_i R_j \right) \\
&= \sum_w \left( \sum_{q \in \mathcal{M}_w} \frac{n_q^Q \left(n_q^Q - 1\right)}{n_q^Q - 1} \left(\pi_q + \gamma_w\right)^2 - \frac{1}{n_w^W - 1} \sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} R_i R_j \right).
\end{aligned}$$

Considering the second term,

$$\begin{aligned}
\sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} R_i R_j &= \sum_{i \in \mathcal{N}_w^W} \sum_{j \in \mathcal{N}_w^W, j \neq i} \left(\pi_{q(i)} + \gamma_w\right)\left(\pi_{q(j)} + \gamma_w\right) \\
&= \sum_{q \in \mathcal{M}_w} \sum_{i \in \mathcal{N}_q} \sum_{j \in \mathcal{N}_w^W, j \neq i} \left(\pi_{q(i)} + \gamma_w\right)\left(\pi_{q(j)} + \gamma_w\right) \\
&= \sum_{q \in \mathcal{M}_w} n_q^Q \left(n_q^Q - 1\right) \left(\pi_q + \gamma_w\right)^2 + \sum_{q \in \mathcal{M}_w} \sum_{i \in \mathcal{N}_q} \sum_{j \in \mathcal{N}_w^W, j \neq i} \left(\pi_{q(i)} + \gamma_w\right)\left(\pi_{q(j)} + \gamma_w\right) \\
&= \sum_{q \in \mathcal{M}_w} n_q^Q \left(n_q^Q - 1\right) \left(\pi_q + \gamma_w\right)^2 + \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma_w\right)\left(\pi_{q'} + \gamma_w\right).
\end{aligned}$$

Since

$$\sum_{q \in \mathcal{M}_w} n_q^Q \left(\pi_q + \gamma_w\right)^2 - \frac{1}{n_w^W - 1} \sum_{q \in \mathcal{M}_w} n_q^Q \left(n_q^Q - 1\right)\left(\pi_q + \gamma_w\right)^2 = \sum_{q \in \mathcal{M}_w} n_q^Q \left(\frac{n_w^W - n_q^Q}{n_w^W - 1}\right)\left(\pi_q + \gamma_w\right)^2,$$

and $n_w^W = \sum_{q \in \mathcal{M}_w} n_q^Q$, we obtain

$$\begin{aligned}
\mu_3 &= \sum_w \left( \sum_{q \in \mathcal{M}_w} n_q^Q \left(\frac{n_w^W - n_q^Q}{n_w^W - 1}\right)\left(\pi_q + \gamma_w\right)^2 - \frac{1}{n_w^W - 1} \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma_w\right)\left(\pi_{q'} + \gamma_w\right) \right) \\
&= \sum_w \frac{1}{n_w^W - 1} \left( \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma_w\right)^2 - \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma_w\right)\left(\pi_{q'} + \gamma_w\right) \right) \\
&= \sum_w \frac{1}{n_w^W - 1} \left( \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma_w\right)\left(\pi_q - \pi_{q'}\right) \right).
\end{aligned}$$

Then, observe that

$$\begin{aligned}
&\sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' \neq q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma\right)\left(\pi_q - \pi_{q'}\right) \\
&= \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' < q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma\right)\left(\pi_q - \pi_{q'}\right) + \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' > q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma\right)\left(\pi_q - \pi_{q'}\right) \\
&= \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' < q} n_q^Q n_{q'}^Q \left(\pi_q + \gamma\right)\left(\pi_q - \pi_{q'}\right) - \sum_{q \in \mathcal{M}_w} \sum_{q' \in \mathcal{M}_w, q' < q} n_q^Q n_{q'}^Q \left(\pi_{q'} + \gamma\right)\left(\pi_q - \pi_{q'}\right)
\end{aligned}$$

$$= \sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'<q} n_q^Q n_{q'}^Q \left(\pi_q - \pi_{q'}\right)\left(\pi_q + \gamma - \pi_{q'} - \gamma\right) = \sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'<q} n_q^Q n_{q'}^Q \left(\pi_q - \pi_{q'}\right)^2,$$

where the second equality switches the indices of $q$ and $q'$ in the second element. Hence,

$$\mu_3 = \sum_w \frac{1}{n_w^W - 1}\left(\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'<q} n_q^Q n_{q'}^Q \left(\pi_q - \pi_{q'}\right)^2\right) \geq 0.$$

Analogously,

$$\begin{aligned}
\mu_2 &= \sum_w \left(\sum_{q\in\mathcal{M}_w}\frac{1}{n_q^Q - 1}\sum_{i\in\mathcal{N}_q^Q}\sum_{j\in\mathcal{N}_q^Q,j\neq i} R_i R_{Yj} - \frac{1}{n_w^W - 1}\sum_{i\in\mathcal{N}_w^W}\sum_{j\in\mathcal{N}_w^W,j\neq i} R_i R_{Yj}\right)\\
&= \sum_w \left(\sum_{q\in\mathcal{M}_w} n_q^Q\left(\pi_q + \gamma_w\right)\left(\pi_{Yq} + \gamma_w\right) - \frac{1}{n_w^W - 1}\sum_{q\in\mathcal{M}_w} n_q^Q\left(n_q^Q - 1\right)\left(\pi_q + \gamma_w\right)\left(\pi_{Yq} + \gamma_w\right)\right)\\
&\quad - \sum_w \frac{1}{n_w^W - 1}\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'\neq q} n_q^Q n_{q'}^Q\left(\pi_q + \gamma_w\right)\left(\pi_{Yq'} + \gamma_w\right)\\
&= \sum_w \left(\sum_{q\in\mathcal{M}_w} n_q^Q\left(\frac{n_w^W - n_q^Q}{n_w^W - 1}\right)\left(\pi_q + \gamma_w\right)\left(\pi_{Yq} + \gamma_w\right) - \frac{1}{n_w^W - 1}\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'\neq q} n_q^Q n_{q'}^Q\left(\pi_q + \gamma_w\right)\left(\pi_{Yq'} + \gamma_w\right)\right)\\
&= \sum_w \frac{1}{n_w^W - 1}\left(\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'\neq q} n_q^Q n_{q'}^Q\left(\pi_q + \gamma_w\right)\left(\pi_{Yq} - \pi_{Yq'}\right)\right)\\
&= \sum_w \frac{1}{n_w^W - 1}\left(\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'<q} n_q^Q n_{q'}^Q\left(\pi_q - \pi_{q'}\right)\left(\pi_{Yq} - \pi_{Yq'}\right)\right),
\end{aligned}$$

and

$$\mu_1 = \sum_w \frac{1}{n_w^W - 1}\left(\sum_{q\in\mathcal{M}_w}\sum_{q'\in\mathcal{M}_w,q'<q} n_q^Q n_{q'}^Q \left(\pi_{Yq} - \pi_{Yq'}\right)^2\right) \geq 0.$$

□

### E.2 Proofs for Appendix A.2

*Proof of Lemma 3.* Suppose not. Then, for some real $\beta_0$,

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

Solving for $\beta_0$,

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

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:

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

Using Table 7, $\sum_k \pi_k^2 = \frac{5}{8}s^2K$, $\sum_k \pi_{Yk}^2 = \left(\frac{5}{8}s^2\beta^2 + h^2\right)K$, and $\sum_k \pi_k\pi_{Yk} = \frac{5}{8}s^2\beta K$, we obtain

$$\frac{1}{4}D = \left(\frac{5}{8}s^2\beta K\right)^2 - \left(\frac{5}{8}s^2K\right)\left(\frac{5}{8}s^2\beta^2 + h^2\right)K = -\frac{5}{8}s^2h^2K^2 \leq 0.$$

Since $h \neq 0$ and $Ks^2 > 0$, there are no real roots of $\beta_0$, a contradiction. □

## E.3 Proofs for Appendix A.3

*Proof of Lemma 4.* I work out the $\mu$'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:

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

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

$$\begin{aligned}
K\sigma_{22} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} \left(G_{ji}G_{ki}E\left[\zeta_i^2\right]R_jR_k + 2G_{ij}G_{ki}E\left[\eta_i\zeta_i\right]R_{Yj}R_k + G_{ij}G_{ik}E\left[\eta_i^2\right]R_{Yj}R_{Yk}\right) \\
&\quad + \sum_i \sum_{j\neq i} \left(G_{ij}^2E\left[\eta_i^2\right]E\left[\zeta_j^2\right] + G_{ij}G_{ji}E\left[\eta_i\zeta_i\right]E\left[\eta_j\zeta_j\right]\right); \\
K\sigma_{11} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} E\left[\zeta_i^2\right]R_{Yj}R_{Yk}\left(G_{ji}G_{ki} + 2G_{ij}G_{ki} + G_{ij}G_{ik}\right) + \sum_i \sum_{j\neq i} E\left[\zeta_i^2\right]E\left[\zeta_j^2\right]\left(G_{ij}^2 + G_{ij}G_{ji}\right); \\
K\sigma_{33} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} E\left[\eta_i^2\right]R_jR_k\left(G_{ji}G_{ki} + 2G_{ij}G_{ki} + G_{ij}G_{ik}\right) + \sum_i \sum_{j\neq i} E\left[\eta_i^2\right]E\left[\eta_j^2\right]\left(G_{ij}^2 + G_{ij}G_{ji}\right); \\
K\sigma_{12} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} \left(G_{ji}G_{ki}E\left[\zeta_i^2\right]R_jR_{Yk} + 2G_{ij}G_{ki}E\left[\zeta_i^2\right]R_{Yj}R_k + G_{ij}G_{ik}E\left[\eta_i\zeta_i\right]R_{Yj}R_{Yk}\right) \\
&\quad + \sum_i \sum_{j\neq i} E\left[\eta_i\zeta_i\right]E\left[\zeta_j^2\right]\left(G_{ij}^2 + G_{ij}G_{ji}\right); \\
K\sigma_{23} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} \left(G_{ji}G_{ki}E\left[\eta_i^2\right]R_{Yj}R_k + 2G_{ij}G_{ki}E\left[\eta_i^2\right]R_jR_{Yk} + G_{ij}G_{ik}E\left[\eta_i\zeta_i\right]R_jR_k\right) \\
&\quad + \sum_i \sum_{j\neq i} E\left[\eta_i\zeta_i\right]E\left[\eta_j^2\right]\left(G_{ij}^2 + G_{ij}G_{ji}\right);\ \text{and} \\
K\sigma_{13} &= \sum_i \sum_{j\neq i} \sum_{k\neq i} E\left[\eta_i\zeta_i\right]R_{Yj}R_k\left(G_{ji}G_{ki} + 2G_{ij}G_{ki} + G_{ij}G_{ik}\right) + \sum_i \sum_{j\neq i} E\left[\eta_i\zeta_i\right]E\left[\eta_j\zeta_j\right]\left(G_{ij}^2 + G_{ij}G_{ji}\right).
\end{aligned}$$

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,

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

Evaluate the terms in the expression. For higher moments of $\pi_k$, $\sum_k \pi_k^2 = Ks^2$, $\sum_k \pi_k^3 = 0$, and $\sum_k \pi_k^4 = Ks^4$. 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,

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

$$\sum_k \omega_{\zeta\eta k}\pi_k\pi_{Yk} = \sum_k \left(\pi_k\sigma_{\xi vk} + \sigma_{vv}\beta + \sigma_{\varepsilon v}\right)\pi_k\left(\pi_k\beta + \sigma_{\xi vk}\right)$$
$$= \sum_k \left(\sigma_{vv}\beta^2\pi_k^2 + \sigma_{\varepsilon v}\pi_k^2\beta + \pi_k^2\sigma_{\xi vk}^2\right) = s^2K\left(\sigma_{vv}\beta^2 + \sigma_{\varepsilon v}\beta + h^2\right).$$

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

$$\sum_k \omega_{\eta\eta k}\omega_{\zeta\zeta k} = \sum_k \sigma_{vv}\left(\pi_k^2\sigma_{\xi\xi} + 2\pi_k\beta\sigma_{\xi vk} + 2\pi_k\sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} - \sigma_{\xi vk}^2 + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v}\beta\right)$$
$$= \sum_k \sigma_{vv}\left(\pi_k^2\sigma_{\xi\xi} + \sigma_{\varepsilon\varepsilon} - \sigma_{\xi vk}^2 + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\xi vk}^2 + 2\sigma_{\varepsilon v}\beta\right)$$
$$= K\sigma_{vv}\left(s^2\sigma_{\xi\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + h^2 + 2\sigma_{\varepsilon v}\beta\right);\ \text{and}$$
$$\sum_k \omega_{\zeta\eta k}^2 = \sum_k \left(\pi_k\sigma_{\xi vk}\pi_k\sigma_{\xi vk} + \sigma_{vv}\beta\pi_k\sigma_{\xi vk} + \sigma_{\varepsilon v}\pi_k\sigma_{\xi vk} + \pi_k\sigma_{\xi vk}\sigma_{vv}\beta + \sigma_{vv}\beta\sigma_{vv}\beta + \sigma_{\varepsilon v}\sigma_{vv}\beta\right)$$
$$+ \sum_k \left(\pi_k\sigma_{\xi vk}\sigma_{\varepsilon v} + \sigma_{vv}\beta\sigma_{\varepsilon v} + \sigma_{\varepsilon v}^2\right)$$
$$= \sum_k \left(\pi_k^2\sigma_{\xi vk}^2 + \sigma_{vv}^2\beta^2 + \sigma_{\varepsilon v}\sigma_{vv}\beta + \sigma_{vv}\beta\sigma_{\varepsilon v} + \sigma_{\varepsilon v}^2\right) = K\left(s^2h^2 + \left(\sigma_{vv}\beta + \sigma_{\varepsilon v}\right)^2\right).$$

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

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

Next, evaluate a few more sums that feature in the other $\sigma$ expressions:

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

Using these results,

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

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

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

Finally, the $\sigma_{11}$ can be obtained:

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

□

*Proof of Proposition 5.* 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 $\mu_1$, observe that:

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

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 $\sigma_{\varepsilon v}=\frac{1}{\sigma_{vv}}\left(\frac{\sigma_{23}c}{2(c-1)}-\sigma_{vv}^2\beta\right)$, $\sigma_{23}=2\frac{c-1}{c}\sigma_{vv}\left(\sigma_{vv}\beta+\sigma_{\varepsilon v}\right)$. From these two expressions, we can observe that:

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

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

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

where the final step uses $\sigma_{\xi\xi} = h/\sigma_{vv}$. This expression for $\sigma_{22}$ is of the form required in Lemma 4. Then,

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

An analogous argument holds for $\sigma_{\xi vk} = -h$. From the $\sigma_{22}$ 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 $det\left(\Sigma_{SF}\right)/h \to C_D \geq 0$. □

## E.4 Derivations for Appendix A.4

**Derivation for continuous setup without covariates**.

This subsection derives expressions for objects in the reduced-form model. Comparing the first-stage equations, $\eta_i = v_i$. As a corollary, for all $i$, $E\left[\eta_i^2\right] = \sigma_{vv}$. Then, $\zeta_i = Z_i'\left(\pi\beta_i - \pi_Y\right) + v_i\beta_i + \varepsilon_i$. Define $\pi_Y$ using $E\left[\zeta_i\right] = 0$ and $E\left[v_i\beta_i\right] = E\left[v_i\left(\beta + \xi_i\right)\right] = \sigma_{\xi vk(i)}$, which implies $\pi_{Yk} = \pi_k\beta + \sigma_{\xi vk}$. Hence, we can rewrite $\zeta_i$ as:

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

By substituting the expression for $\zeta_i$, the covariance is $E\left[\eta_i\zeta_i \mid k\right] = \pi_k\sigma_{\xi vk} + \sigma_{vv}\beta + E\left[v_i^2\xi_i\right] + \sigma_{\varepsilon v}$. By Isserlis' theorem, $E\left[v_i^2\xi_i\right] = 0$, so $E\left[\eta_i\zeta_i \mid k\right] = \pi_k\sigma_{\xi vk} + \sigma_{vv}\beta + \sigma_{\varepsilon v}$. The variance of $\zeta_i$ can be derived analogously. Since $E\left[v_i^2\beta_i^2\right] = \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\xi vk}^2$ by applying Isserlis' theorem, with $\omega_{\eta\eta k} := E[\eta_i^2 \mid k(i) = k]$, $\omega_{\zeta\eta k} := E[\zeta_i\eta_i \mid k(i) = k]$, and $\omega_{\zeta\zeta k} := E[\zeta_i^2 \mid k(i) = k]$, we obtain:

$$\begin{aligned}
\omega_{\eta\eta k} &= \sigma_{vv}^2, \\
\omega_{\zeta\eta k} &= \pi_k\sigma_{\xi vk} + \sigma_{vv}\beta + \sigma_{\varepsilon v}, \text{ and} \\
\omega_{\zeta\zeta k} &= \pi_k^2\sigma_{\xi\xi} + 2\pi_k\beta\sigma_{\xi vk} + 2\pi_k\sigma_{\varepsilon\xi} + \sigma_{\varepsilon\varepsilon} + \sigma_{\xi vk}^2 + \sigma_{vv}\beta^2 + \sigma_{vv}\sigma_{\xi\xi} + 2\sigma_{\varepsilon v}\beta.
\end{aligned} \tag{24}$$

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

$$LATE_k = \frac{\pi_{Yk}}{\pi_k} = \beta + \frac{\sigma_{\xi vk}}{\pi_k}. \tag{25}$$

**Derivation for binary setup without covariates**.

The reduced-form residuals are given by:

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

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

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

Due to the judge setup, the estimand is:

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

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

**Derivation for binary setup with covariates**.

Consider the structural model:

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

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

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

Using the result that $\pi_{Yk} = \pi_k\beta - (1-s)(2p-1)\sigma_{\xi vk}$,

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

Analogously, $\sum_{i\in\mathcal{N}_t}\sum_{j\in\mathcal{N}_t\setminus i} G_{ij}R_iR_j = \frac{25}{9}\pi_t^2$. 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 $\beta$ as our estimand:

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

regardless of $\gamma_t$.