Supplemental Appendix for "Inference for Two-Stage Extremum Estimators"

Aristide Houndetoungan[*] and Abdoul Haki Maoude[†]

February 6, 2024

Abstract

This supplemental appendix includes additional results and technical details omitted from the main text. Section S.1 discusses primitive conditions of some high-level assumptions introduced in the paper. Section S.2 provides technical details regarding our simulation study, including a description of how we approximate the asymptotic distributions. Section S.3 presents detailed results of our empirical application with network data.

S.1 Online Appendix–Proofs

S.1.1 Consistency of Plug-in Estimators

We impose lower-level assumptions that result in a consistent plug-in estimator. For notational ease, we omit $\mathbf{y}_n$ and $\mathbf{X}_n$ in $Q_n(\boldsymbol{\theta},\ \mathbf{y}_n,\ \mathbf{X}_n,\ \mathbf{B})$ and we simply write $Q_n(\boldsymbol{\theta},\ \mathbf{B})$. We also write $q_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_i)$ instead of $q(\boldsymbol{\theta},\ y_i,\ \boldsymbol{x}_i,\ \boldsymbol{\beta}_i)$. For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $a, \tau > 0$, we define

$$\mathcal{H}_i(a,\tau) = \sup_{\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\| < \tau} \frac{|q_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_i) - q_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_{0,i})|}{\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|^a},$$

where $\|.\|$ is the ℓ^2-norm.

Assumption S.1 (Primitive Conditions for Assumption 2.3)**.**

(i) $\boldsymbol{\Theta}$ *is a compact subset and* $\boldsymbol{\theta}_0$ *is an interior point of* $\boldsymbol{\Theta}$*.*

(ii) The function $Q_n(\boldsymbol{\theta},\ \mathbf{B}_0)$ *converges uniformly in probability (across* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*) to a nonstochastic function* $Q_0(\boldsymbol{\theta})$ *that is maximized only at* $\boldsymbol{\theta}_0$*.*

[*]Cy Cergy Paris University and Thema – Corresponding author – Email address: aristide.houndetoungan@cyu.f
[†]Concordia University, Department of Statistics – Email address: abdoulhaki.maoude@concordia.ca

(iii) For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, there exists constants $a(\boldsymbol{\theta})$, $\tau(\boldsymbol{\theta}) > 0$, such that, $\max_i \mathcal{H}_i(a(\boldsymbol{\theta}), \tau(\boldsymbol{\theta})) = O_p(1)$.

(iv) There exists a sequence of neighborhoods $\mathcal{O}(\boldsymbol{\beta}_{0,1}), \ \ldots, \ \mathcal{O}(\boldsymbol{\beta}_{0,n})$, such that $\partial_{\boldsymbol{\theta}} q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i)$ is $O_p(1)$, uniformly across $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $\boldsymbol{\beta}_i \in \mathcal{O}(\boldsymbol{\beta}_{0,i})$ and i, that is, $\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \ \boldsymbol{\beta}_i \in \mathcal{O}(\boldsymbol{\beta}_{0,i}), \ i \leqslant n} \|\partial_{\boldsymbol{\theta}} q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i)\| = O_p(1)$.

The compactness restriction in Condition (i) allows for the plug-in estimator to converge in its support. Condition (ii) is a classical identification condition also required for a standard M-estimator. Importantly, this condition does not involve the estimator $\hat{\mathbf{B}}_n$. It is an identification condition that is set at the true $\mathbf{B}_0$. Condition (iii) implies that $|q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i) - q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_{0,i})| \leqslant \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|^{a(\boldsymbol{\theta})} O_p(1)$ for all i and $\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\| < \tau(\boldsymbol{\theta})$. A similar assumption is also imposed by Cattaneo et al. (2019) and requires $q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i)$ to be smooth in $\boldsymbol{\beta}_i$. We use this condition and the uniform convergence of $\hat{\boldsymbol{\beta}}_{n,i}$ (Assumption 2.1) to show that $Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)$ converges in probability to 0 for each $\boldsymbol{\theta}$. Condition (iv) allows us to generalize this point-wise convergence to a uniform convergence.[1]

Proposition S.1. *Under Assumptions 2.1, 2.2, and S.1, the estimator $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\boldsymbol{\theta}_0$.*

Proof. The proof is performed in two steps.

Step 1: We show that $Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n)$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$.
For any $\boldsymbol{\theta}$, we have $|Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_0(\boldsymbol{\theta})| \leqslant |Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)| + |Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0) - Q_0(\boldsymbol{\theta})|$. Since $Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0) - Q_0(\boldsymbol{\theta})$ converges uniformly in probability in $\boldsymbol{\theta}$ to 0 (Condition (ii) of Assumption S.1), it is sufficient to show that $Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)$ also converges uniformly in probability in $\boldsymbol{\theta}$ to 0.

By Assumption 2.1, for n large enough, $\max_i \|\hat{\boldsymbol{\beta}}_{n,i} - \boldsymbol{\beta}_{0,i}\| < \tau(\boldsymbol{\theta})$ with probability approaching one. Thus, by Condition (iii) of Assumption S.1,

$$|q_i(\boldsymbol{\theta}, \ \hat{\boldsymbol{\beta}}_{n,i}) - q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_{0,i})| \leqslant \|\hat{\boldsymbol{\beta}}_{n,i} - \boldsymbol{\beta}_{0,i}\|^{a(\boldsymbol{\theta})} \max_i \mathcal{H}_i(a(\boldsymbol{\theta}), \tau(\boldsymbol{\theta})) \quad \text{for all } i$$

with probability approaching one. As $|Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)| \leqslant \max_i |q_i(\boldsymbol{\theta}, \ \hat{\boldsymbol{\beta}}_{n,i}) - q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_{0,i})|$, this implies that

$$|Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)| \leqslant \max_i \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|^{a(\boldsymbol{\theta})} \max_i \mathcal{H}_i(a(\boldsymbol{\theta}), \tau(\boldsymbol{\theta}))$$

with probability approaching one. As a result $Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)$ converges in probability to zero because $\max_i \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|^{a(\boldsymbol{\theta})} = o_p(1)$ and $\max_i \mathcal{H}_i(a(\boldsymbol{\theta}), \tau(\boldsymbol{\theta})) = O_p(1)$.

To show that the convergence is uniform, we apply the mean value theorem to $Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0)$ with respect to $\boldsymbol{\theta}$. For any $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, we have

$$Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0) - \big(Q_n(\tilde{\boldsymbol{\theta}}, \ \hat{\mathbf{B}}_n) - Q_n(\tilde{\boldsymbol{\theta}}, \ \mathbf{B}_0)\big) = (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' \hat{\boldsymbol{Q}}_n,$$

where $\hat{\boldsymbol{Q}}_n = \frac{1}{n} \sum_{i=1}^n \partial_{\boldsymbol{\theta}} \big(q_i(\boldsymbol{\theta}_n^+, \ \hat{\boldsymbol{\beta}}_{n,i}) - q_i(\boldsymbol{\theta}_n^+, \ \boldsymbol{\beta}_{0,i})\big)$, for some $\boldsymbol{\theta}_n^+$ that lies between $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$. Thus,

$$|Q_n(\boldsymbol{\theta}, \ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta}, \ \mathbf{B}_0) - \big(Q_n(\tilde{\boldsymbol{\theta}}, \ \hat{\mathbf{B}}_n) - Q_n(\tilde{\boldsymbol{\theta}}, \ \mathbf{B}_0)\big)| \leqslant \|\hat{\boldsymbol{Q}}_n\| \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|.$$

[1] Note that $\partial_{\boldsymbol{\theta}} q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i)$ is be bounded in probability for each i as it must have finite moments. Consequently, one can find a neighborhood $\mathcal{O}(\boldsymbol{\beta}_{0,i})$ such $\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, \ \boldsymbol{\beta}_i \in \mathcal{O}(\boldsymbol{\beta}_{0,i})} \|\partial_{\boldsymbol{\theta}} q_i(\boldsymbol{\theta}, \ \boldsymbol{\beta}_i)\|$ is bounded in probability. The additional requirement of Condition (iv) is that the bound is uniform across i as n grows to infinity.

As $\hat{\boldsymbol{Q}}_n = O_p(1)$ (Condition (iv) of Assumption S.1) and $\boldsymbol{\Theta}$ is compact, it follows from Lemma 2.9 of Newey and McFadden (1994) that $Q_n(\boldsymbol{\theta},\ \hat{\mathbf{B}}_n) - Q_n(\boldsymbol{\theta},\ \mathbf{B}_0)$ converges uniformly in probability to 0.

Step 2: We establish the consistency of the estimator $\hat{\boldsymbol{\theta}}_n$.

Let $\mathcal{O}(\boldsymbol{\theta}_0)^c$ be the complement of $\mathcal{O}(\boldsymbol{\theta}_0)$ in $\boldsymbol{\Theta}$. Note that $\mathcal{O}(\boldsymbol{\theta}_0)^c$ is nonempty and compact and $\max_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_0)^c} Q_0(\boldsymbol{\theta})$ exists. Let $\delta = Q_0(\boldsymbol{\theta}_0) - \max_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_0)^c} Q_0(\boldsymbol{\theta})$ and $J_n = \big\{|Q_n(\boldsymbol{\theta},\ \hat{\mathbf{B}}_n) - Q_0(\boldsymbol{\theta})| < \delta/2,\ \text{ for all } \boldsymbol{\theta}\big\}$. We know that $Q_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$ and that $\boldsymbol{\theta}_0 \notin \mathcal{O}(\boldsymbol{\theta}_0)^c$. Thus $\delta > 0$. Moreover, since $Q_n(\boldsymbol{\theta},\ \hat{\mathbf{B}}_n)$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$, we have $\lim \mathbb{P}(J_n) = 1$.

$$J_n \implies \big\{Q_0(\hat{\boldsymbol{\theta}}_n) > Q_n(\hat{\boldsymbol{\theta}}_n, \hat{\mathbf{B}}_n) - \delta/2\big\} \cap \big\{Q_n(\boldsymbol{\theta}_0,\ \hat{\mathbf{B}}_n) > Q_0(\boldsymbol{\theta}_0) - \delta/2\big\} \tag{S.1}$$

As $\hat{\boldsymbol{\theta}}_n = \operatorname{argmax}_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta},\ \hat{\mathbf{B}}_n)$, we also have $Q_n(\hat{\boldsymbol{\theta}}_n,\ \hat{\mathbf{B}}_n) \geqslant Q_n(\boldsymbol{\theta}_0,\ \hat{\mathbf{B}}_n)$. Thus, $\big\{Q_0(\hat{\boldsymbol{\theta}}_n) > Q_n(\hat{\boldsymbol{\theta}}_n, \hat{\mathbf{B}}_n) - \delta/2\big\}$ implies $\big\{Q_0(\hat{\boldsymbol{\theta}}_n) > Q_n(\boldsymbol{\theta}_0, \hat{\mathbf{B}}_n) - \delta/2\big\}$. It follows from (S.1) that

$$\begin{aligned} J_n &\implies \big\{Q_0(\hat{\boldsymbol{\theta}}_n) > Q_n(\boldsymbol{\theta}_0, \hat{\mathbf{B}}_n) - \delta/2\big\} \cap \big\{Q_n(\boldsymbol{\theta}_0,\ \hat{\mathbf{B}}_n) > Q_0(\boldsymbol{\theta}_0) - \delta/2\big\}, \\ J_n &\implies \big\{Q_0(\hat{\boldsymbol{\theta}}_n) > Q_0(\boldsymbol{\theta}_0) - \delta\big\}. \end{aligned} \tag{S.2}$$

As $\delta = Q_0(\boldsymbol{\theta}_0) - \max_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_0)^c} Q_0(\boldsymbol{\theta})$, it turns out from (S.2) that

$$\begin{aligned} J_n &\implies \big\{Q_0(\hat{\boldsymbol{\theta}}_n) > \max_{\boldsymbol{\theta}\in\mathcal{O}(\boldsymbol{\theta}_0)^c} Q_0(\boldsymbol{\theta})\big\}, \\ J_n &\implies \hat{\boldsymbol{\theta}}_n \in \mathcal{O}(\boldsymbol{\theta}_0). \end{aligned} \tag{S.3}$$

As $\lim \mathbb{P}(J_n) = 1$, then (S.3) implies that $\lim \mathbb{P}(\hat{\boldsymbol{\theta}}_n \in \mathcal{O}(\boldsymbol{\theta}_0)) = 1$. This is true for any open subset $\mathcal{O}(\boldsymbol{\theta}_0)$ that contains $\boldsymbol{\theta}_0$. As a result, $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\boldsymbol{\theta}_0$. □

S.1.2 Primitive Conditions for Assumption 4.2

For notational ease, let $\ddot{q}_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_i) = \partial_{\boldsymbol{\theta}}\partial_{\boldsymbol{\theta}'} q(\boldsymbol{\theta},\ y_i,\ \boldsymbol{x}_i,\ \boldsymbol{\beta}_i)$. For any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $a, \tau\ > 0$, we define

$$\ddot{\mathcal{H}}_i(a,\tau) = \sup_{(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|)^a < \tau} \frac{\|\ddot{q}_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_i) - \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\|}{(\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|)^a}.$$

We impose lower-level conditions that imply Assumption 4.2.

Assumption S.2 (Primitive Conditions for Assumption 4.2)**.**

(i) The matrix $\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})$ *converges in probability to a finite nonsingular matrix* $\mathbf{A}_0$ *defined by* $\lim \mathbb{E}\big(\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\big)$.

(ii) There exists constants $a^*,\ \tau^* > 0$ *such that* $\max_i \ddot{\mathcal{H}}_i(a^*, \tau^*) = O_p(1)$.

Condition (i) imposes that $\frac{1}{n}\sum_i^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})$ converges in to $\lim \mathbb{E}\big(\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\big)$. This condition is classical as in the case of a single-step estimator. It does not involve any estimator and can be implied by the weak law of large numbers (WLLN). For $\ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})$'s dependent across i, WLLN for dependent processes can be used. Condition (ii) is similar to Condition (iii) of Assumption S.1. It requires $\ddot{q}_i(\boldsymbol{\theta},\ \boldsymbol{\beta}_i)$ to be smooth in both $\boldsymbol{\beta}_i$ and $\boldsymbol{\theta}$, uniformly in i.

Proposition S.2. *Under Assumptions 2.1, 2.2, 2.3, S.1 and S.2, the Hessian of the objective function evaluated at any consistent estimator $\boldsymbol{\theta}_n^+$, given by $\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i})$, converges in probability to a finite nonsingular matrix $\mathbf{A}_0 = \lim \mathbb{E}\big(\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\big)$.*

Proof. By Assumption 2.1, for n large enough, $\|\boldsymbol{\theta}_n^+ - \boldsymbol{\theta}_0\| + \max_i \|\hat{\boldsymbol{\beta}}_{n,i} - \boldsymbol{\beta}_{0,i}\| < \tau^*$ with probability approaching one. Thus, $\|\ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i}) - \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\| \leqslant (\|\boldsymbol{\theta}_n^+ - \boldsymbol{\theta}_0\| + \|\hat{\boldsymbol{\beta}}_{n,i} - \boldsymbol{\beta}_{0,i}\|)^{a^*} \max_i \ddot{\mathcal{H}}_i(a^*, \tau^*)$, for all i, with probability approaching one.

As $\|\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i}) - \frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\| \leqslant \max_i \|\ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i}) - \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\|$, it follows that $\|\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i}) - \frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})\| \leqslant (\|\boldsymbol{\theta}_n^+ - \boldsymbol{\theta}_0\| + \max_i \|\hat{\boldsymbol{\beta}}_{n,i} - \boldsymbol{\beta}_{0,i}\|)^{a^*} \max_i \ddot{\mathcal{H}}_i(a^*, \tau^*)$ with probability approaching one, where $\|\boldsymbol{\theta}_n^+ - \boldsymbol{\theta}_0\| + \max_i \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_{0,i}\|^{a(\boldsymbol{\theta})} = o_p(1)$ and $\max_i \ddot{\mathcal{H}}_i(a^*, \tau^*) = O_p(1)$. As a result, $\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_n^+,\ \hat{\boldsymbol{\beta}}_{n,i}) - \frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i}) = o_p(1)$. Given that $\frac{1}{n}\sum_{i=1}^n \ddot{q}_i(\boldsymbol{\theta}_0,\ \boldsymbol{\beta}_{0,i})$ converges in probability to a finite nonsingular matrix $\mathbf{A}_0$, the result follows. □

S.2 Supplementary Materials on the Simulation Study

This section provides a detailed explanation of the method used to estimate the asymptotic variance and the asymptotic cumulative distribution function (CDF) of the plug-in estimators in the simulation study. Our replication code available at https://github.com/ahoundetoungan/InferenceTSE implements this method. We also present the estimates of the asymptotic CDF of the debiased estimators.

S.2.1 Asymptotic Variance and Asymptotic CDF

DGPs A and B

DGP A is a treatment effect model with endogeneity. The model is defined as follows:

$$y_i = \theta_0 d_i + \varepsilon_i, \quad d_i = \mathbb{1}\{z_i > 0.5(\varepsilon_i + 1.2)\}, \quad z_i \sim \text{Uniform}[0,\ 1], \quad \varepsilon_i \sim \text{Uniform}[-1, 1],$$

where d_i is a treatment status indicator, z_i is an instrument for the treatment and $\theta_0 = 1$. In the first stage, we perform two OLS regressions: a regression of y_i on $\boldsymbol{z}_i = (1, z_i)'$ and another regression of d_i on $\boldsymbol{z}_i$. For DGP B, the vector of regressors in the first stage is $\boldsymbol{z}_i = (1,\ z_{1,i},\ \dots,\ z_{k_n,i})'$.

Let $\hat{\boldsymbol{\gamma}}_n^{(y)}$ and $\hat{\boldsymbol{\gamma}}_n^{(d)}$ be the respective OLS estimators and let $\hat{\boldsymbol{\gamma}}_n = (\hat{\boldsymbol{\gamma}}_n^{(y)\prime},\ \hat{\boldsymbol{\gamma}}_n^{(d)\prime})'$ be the joint first-stage estimator. Let also $\hat{\nu}_i^{(y)}$ and $\hat{\nu}_i^{(d)}$ be the residuals of the regressions; that is, $\hat{\nu}_i^{(y)} = y_i - \boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(y)}$ and $\hat{\nu}_i^{(d)} = d_i - \boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)}$. We define $\boldsymbol{z}_i^{(\nu)} = (\hat{\nu}_i^{(y)}\boldsymbol{z}_i',\ \hat{\nu}_i^{(d)}\boldsymbol{z}_i')'$. The estimator of the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_n$ is a normal distribution with mean $\hat{\boldsymbol{\gamma}}_n$ and covariance matrix $\hat{\mathbb{V}}(\hat{\boldsymbol{\gamma}}_n) = \frac{1}{n}\hat{\mathbf{H}}_n^{-1}\hat{\mathbf{J}}_n\hat{\mathbf{H}}_n^{-1}$, where

$$\hat{\mathbf{H}}_n = \frac{1}{n}\sum_{i=1}^n \operatorname{diag}(\textstyle\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i',\ \sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i') \quad \text{and} \quad \hat{\mathbf{J}}_n = \frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i^{(\nu)}\boldsymbol{z}_i^{(\nu)\prime}.$$

The notation diag stands for the bloc diagonal matrix operator. In the second stage, the objective function to be maximized is $Q_n(\theta,\ \mathbf{y}_n,\ \hat{\mathbf{B}}_n) = -\frac{1}{n}\sum_{i=1}^n(\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(y)} - \theta\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)})^2$, where $\hat{\mathbf{B}}_n = (\hat{\boldsymbol{\beta}}_{n,1},\ \dots,\ \hat{\boldsymbol{\beta}}_{n,n})'$

and $\hat{\boldsymbol{\beta}}_{n,i} = (\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(y)},\ \boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)})'$. This implies that $\dot{\boldsymbol{q}}_n(\mathbf{y}_n,\ \hat{\mathbf{B}}_n) = \frac{2}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)}(\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(y)} - \theta_0\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)})$. We define the following expressions:

$$\hat{A}_n = \frac{1}{n}\sum_{i=1}^n(\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d)})^2 \quad \text{and} \quad \hat{\mathcal{E}}_{n,s} = \frac{2}{\sqrt{n}}\sum_{i=1}^n \boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d,s)}(\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(y,s)} - \hat{\theta}_n\boldsymbol{z}_i'\hat{\boldsymbol{\gamma}}_n^{(d,s)}),$$

where $(\hat{\boldsymbol{\gamma}}_n^{(y,1)\prime},\ \hat{\boldsymbol{\gamma}}_n^{(d,1)\prime})',\ \dots,\ (\hat{\boldsymbol{\gamma}}_n^{(y,\kappa)\prime},\ \hat{\boldsymbol{\gamma}}_n^{(d,\kappa)\prime})' \overset{i.i.d}{\sim} N(\hat{\boldsymbol{\gamma}}_n,\ \hat{\mathbb{V}}(\hat{\boldsymbol{\gamma}}_n))$. Let $\hat{\psi}_{n,s} = \frac{\hat{\mathcal{E}}_{n,s}}{\hat{A}_n}$. The asymptotic CDF of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be estimated by the empirical CDF of the sample: $\{\hat{\psi}_{n,s},\ s = 1,\ \dots,\ \kappa\}$. The estimator of the asymptotic variance of $\hat{\theta}_n$ is

$$\hat{\mathbb{V}}(\hat{\theta}_n) = \frac{\sum_{s=1}^{\kappa}(\hat{\mathcal{E}}_{n,s} - \hat{\mathbb{E}}(\mathcal{E}_n))^2}{n(\kappa-1)\hat{A}_n^2},$$

where $\hat{\mathbb{E}}(\mathcal{E}_n) = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathcal{E}}_{n,s}$. The debiased estimator is given by

$$\theta^*_{n,\kappa} = \hat{\theta}_n - \hat{\mathbb{E}}(\mathcal{E}_n)/(\sqrt{n}\hat{A}_n).$$

Let $\hat{\psi}^*_{n,s} = \frac{\hat{\mathcal{E}}^*_{n,s} - \hat{\mathbb{E}}(\mathcal{E}^*_n)}{\hat{A}^*_n}$, where $\hat{A}^*_n$, $\hat{\mathcal{E}}^*_{n,s}$, and $\hat{\mathbb{E}}(\mathcal{E}^*_n)$ are defined as $\hat{A}_n$, $\hat{\mathcal{E}}_{n,s}$, and $\hat{\mathbb{E}}(\mathcal{E}_n)$, respectively, with the difference that they are computed using $\theta^*_{n,\kappa}$ and not $\hat{\theta}_n$. We can estimate the asymptotic CDF of $\sqrt{n}(\theta^*_{n,\kappa} - \theta_0)$ by the empirical CDF of the sample: $\{\hat{\psi}^*_{n,s},\ s = 1,\ \dots,\ \kappa\}$.

DGP C

DGP C is a Poisson model with a latent covariate that is defined as:

$$y_i \sim \text{Poisson}(\exp(\theta_{0,1} + \theta_{0,2}p_i)), \quad p_i = \sin^2(\pi z_i), \quad z_i \sim \text{Uniform}[0,\ 10], \quad d_i \sim \text{Bernoulli}(p_i),$$

where p_i is an unobserved probability and $\boldsymbol{\theta}_0 = (\theta_{0,1},\ \theta_{0,2})' = (0.5,\ 2)'$. The practitioner observes the pairs $(y_i,\ z_i)$ for all i but only observes d_i for a representative subsample of size $n^* \leqslant n$. In the first stage, $p_i = \mathbb{E}(d_i|z_i)$ is estimated using a generalized additive model (GAM) of d_i on z_i in the subsample of size n^* where d_i is observed. The GAM involves approximating p_i by piecewise polynomial functions of z_i.[2] We consider cubic polynomial functions on the intervals $[0,\ 0.5],\ \dots,\ [9.5,\ 10]$. This approach can thus be regarded as an OLS regression of d_i on numerous explanatory variables called bases, which are computed from z_i. We can write $\hat{p}_i = h(z_i,\ \hat{\boldsymbol{\gamma}}_n)$, where h is a piecewise cubic polynomial function and $\hat{\boldsymbol{\gamma}}_n$ is the OLS estimator. The regression results can be used to compute $\hat{p}_i$ for any i in the full sample because we observe z_i of all i. The estimator of the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_n$ is a normal distribution with mean $\hat{\boldsymbol{\gamma}}_n$ and the standard OLS variance denoted $\hat{\mathbb{V}}(\hat{\theta}_n)$.

In the second stage, we perform a maximum likelihood (ML) estimation by assuming that y_i follows a Poisson distribution with mean $\exp(\hat{\boldsymbol{\beta}}_{n,i}'\boldsymbol{\theta})$, where $\hat{\boldsymbol{\beta}}_{n,i} = (1,\ \hat{p}_i)$. The objective function is thus given by $Q_n(\theta_0,\ \mathbf{y}_n,\ \hat{\mathbf{B}}_n) = \frac{1}{n}\sum_{i=1}^n\left(y_i\hat{\boldsymbol{\beta}}_{n,i}'\boldsymbol{\theta} - \exp(\hat{\boldsymbol{\beta}}_{n,i}'\boldsymbol{\theta})\right)$ and $\dot{\boldsymbol{q}}_n(\mathbf{y}_n,\ \hat{\mathbf{B}}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(y_i - \exp(\hat{\boldsymbol{\beta}}_{n,i}'\boldsymbol{\theta}_0)\right)\hat{\boldsymbol{\beta}}_{n,i}$. Therefore,

$$\hat{\mathbf{A}}_n = \frac{1}{n}\sum_{i=1}^n \exp\left(\hat{\boldsymbol{\beta}}_{n,i}'\hat{\boldsymbol{\theta}}_n\right)\hat{\boldsymbol{\beta}}_{n,i}\hat{\boldsymbol{\beta}}_{n,i}', \quad \hat{\mathbf{V}}_n = \frac{1}{n}\sum_{i=1}^n \exp\left(\hat{\boldsymbol{\beta}}_{n,i}'\hat{\boldsymbol{\theta}}_n\right)\hat{\boldsymbol{\beta}}_{n,i}\hat{\boldsymbol{\beta}}_{n,i}', \quad \text{and}$$

$$\hat{\mathcal{E}}_{n,s} = \frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\exp(\hat{\boldsymbol{\beta}}_{n,i}^{(s)\prime}\hat{\boldsymbol{\theta}}_n) - \exp(\hat{\boldsymbol{\beta}}_{n,i}^{(s)\prime}\hat{\boldsymbol{\theta}}_n)\right)\hat{\boldsymbol{\beta}}_{n,i}^{(s)},$$

[2] See HASTIE, T. J. (2017): "Generalized additive models," in *Statistical models in S*, Routledge, 249–307.

where $\hat{\boldsymbol{\beta}}_{n,i}^{(s)} = h(z_i,\ \hat{\boldsymbol{\gamma}}_n^{(s)})$, for $s = 1,\ \ldots,\ \kappa$, and $\hat{\boldsymbol{\gamma}}_n^{(1)}\ \ldots,\ \hat{\boldsymbol{\gamma}}_n^{(\kappa)} \overset{i.i.d}{\sim} N(\hat{\boldsymbol{\gamma}}_n,\ \hat{\mathbb{V}}(\hat{\boldsymbol{\gamma}}_n))$.

Let $\hat{\boldsymbol{\psi}}_{n,s} = \hat{\mathbf{A}}_n^{-1}\hat{\mathbf{V}}_n^{1/2}\boldsymbol{\zeta}_s + \hat{\mathbf{A}}_n^{-1}\hat{\mathcal{E}}_{n,s}$, where $\boldsymbol{\zeta}_1,\ \ldots,\ \boldsymbol{\zeta}_\kappa \overset{i.i.d}{\sim} N(0,\ \boldsymbol{I}_{K_\theta})$. The asymptotic CDF of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ can be estimated by the empirical CDF of the sample: $\{\hat{\boldsymbol{\psi}}_{n,s},\ s = 1,\ \ldots,\ \kappa\}$. The asymptotic variance of $\hat{\boldsymbol{\theta}}_n$ is estimated by

$$\hat{\mathbb{V}}(\hat{\boldsymbol{\theta}}_n) = \frac{\hat{\mathbf{A}}_n^{-1}\hat{\boldsymbol{\Sigma}}_n^{\kappa}\hat{\mathbf{A}}_n^{-1}}{n},$$

where $\hat{\boldsymbol{\Sigma}}_n^{\kappa} = \hat{\mathbf{V}}_n + \frac{1}{\kappa-1}\sum_{s=1}^{\kappa}(\hat{\mathcal{E}}_{n,s} - \hat{\boldsymbol{\Omega}}_n^{\kappa})(\hat{\mathcal{E}}_{n,s} - \hat{\boldsymbol{\Omega}}_n^{\kappa})'$ and $\hat{\boldsymbol{\Omega}}_n^{\kappa} = \frac{1}{\kappa}\sum_{s=1}^{\kappa}\hat{\mathcal{E}}_{n,s}$. The debiased estimator is given by

$$\boldsymbol{\theta}_{n,\kappa}^{*} = \hat{\boldsymbol{\theta}}_n - \hat{\mathbf{A}}_n^{-1}\hat{\boldsymbol{\Omega}}_n^{\kappa}/\sqrt{n}.$$

Let $\hat{\boldsymbol{\psi}}_{n,s}^{*} = (\hat{\mathbf{A}}_n^{*})^{-1}(\hat{\mathbf{V}}_n^{*})^{1/2}\boldsymbol{\zeta}_s + (\hat{\mathbf{A}}_n^{*})^{-1}(\hat{\mathcal{E}}_{n,s}^{*} - \hat{\boldsymbol{\Omega}}_n^{*\kappa})$, where $\hat{\mathbf{A}}_n^{*}$, $\hat{\mathbf{V}}_n^{*}$, $\hat{\mathcal{E}}_{n,s}^{*}$, and $\hat{\boldsymbol{\Omega}}_n^{*\kappa}$ are respectively defined as $\hat{\mathbf{A}}_n$, $\hat{\mathbf{V}}_n$, $\hat{\mathcal{E}}_{n,s}$, and $\hat{\boldsymbol{\Omega}}_n^{\kappa}$, with the difference that they are computed using $\boldsymbol{\theta}_{n,\kappa}^{*}$ and not $\hat{\boldsymbol{\theta}}_n$. We can estimate the asymptotic CDF of $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{*} - \boldsymbol{\theta}_0)$ by the empirical CDF of the sample: $\{\hat{\boldsymbol{\psi}}_{n,s}^{*},\ s = 1,\ \ldots,\ \kappa\}$.

DGP D

DGP D is a copula-based multivariate time-series model. We consider k_n returns $y_{1,i},\ \ldots,\ y_{k_n,i}$, where i is time and $k_n \geqslant 2$. Each $y_{p,i}$, for $p = 2,\ \ldots,\ k_n$, follows an AR(1)-GARCH(1, 1) model, such that

$$y_{p,i} = \phi_{p,0} + \phi_{p,1}y_{p,i-1} + \sigma_{p,i}\varepsilon_{p,i}, \quad \sigma_{p,i}^2 = \beta_{p,0} + \beta_{p,1}\sigma_{p,i-1}^2\varepsilon_{p,i-1}^2 + \beta_{p,2}\sigma_{p,i-1}^2,$$

where $\phi_{p,0} = 0$, $\phi_{p,i-1} = 0.4$, $\beta_{p,0} = 0.05$, $\beta_{p,1} = 0.05$, $\beta_{p,2} = 0.9$, and $\varepsilon_{p,i}$ follows a standardized Student distribution of degree-of-freedom $\nu_p = 6$. We account for the correlation between the returns using the Clayton copula. The joint density function of $y_i = (y_{1,i},\ \ldots,\ y_{p,i})'$ conditional on $\mathcal{F}^{i-1}$ is given by $c_i(G_{1,i}(\boldsymbol{\beta}_{0,1}),\ \ldots,\ G_{k_n,i}(\boldsymbol{\beta}_{0,k_n}),\ \theta_0)$, where $\boldsymbol{\beta}_{0,p} = (\phi_{p,0},\ \phi_{p,1},\ \beta_{p,0},\ \beta_{p,1},\ \beta_{p,2},\ \nu_p)'$, $G_{p,i}(\boldsymbol{\beta}_{0,p})$ is the CDF of $y_{p,i}$ conditional on $\mathcal{F}^{i-1}$, and c_i is the probability density function (PDF) of k_n-dimensional Clayton copula of parameter $\theta_0 = 4$. A multi-stage estimation strategy can be used to estimate θ_0. In the first k_n stages, we separately estimate each $\boldsymbol{\beta}_{0,p}$ by applying an AR(1)-GARCH(1, 1) model to the sample $y_{p,1},\ \ldots,\ y_{p,n}$.

$$\hat{\boldsymbol{\beta}}_{n,p} = \arg\max_{\boldsymbol{\beta}_p} \ell_p := \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\underbrace{\log g_{p,i}(\boldsymbol{\beta}_p)}_{\ell_{p,i}}, \quad \text{for } p = 1,\ \ldots,\ k_n\,,$$

where $g_{p,i}(\boldsymbol{\beta}_{0,p})$ is the PDF of $y_{p,i}$ conditional on $\mathcal{F}^{i-1}$. Let $\hat{\boldsymbol{\beta}}_n$ be the estimator $\boldsymbol{\beta}_0 := (\boldsymbol{\beta}_{0,1}',\ \ldots,\ \boldsymbol{\beta}_{0,k_n}')'$. The estimator of the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ is a normal distribution with mean $\hat{\boldsymbol{\beta}}_n$ and variance given by $\hat{\mathbb{V}}(\hat{\boldsymbol{\beta}}_n) = \frac{1}{n}\hat{\mathbf{H}}_n^{-1}\hat{\mathbf{J}}_n\hat{\mathbf{H}}_n^{-1}$, where

$$\hat{\mathbf{H}}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\operatorname{diag}(\tfrac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}\ell_{1,i},\ \ldots,\ \tfrac{\partial^2}{\partial\boldsymbol{\beta}'\partial\boldsymbol{\beta}}\ell_{k_n,i}) \quad \text{and} \quad \hat{\mathbf{J}}_n = \mathbb{V}_{\text{HAC}}(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^{n}(\partial_{\boldsymbol{\beta}'}\ell_{1,i},\ \ldots,\ \partial_{\boldsymbol{\beta}'}\ell_{k_n,i})')\,.$$

The notation $\mathbb{V}_{\text{HAC}}$ is the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix to account for the serial correlation.[3] In the HAC approach, we use the quadratic spectral kernel and set the bandwidth to $\frac{3}{4}n^{1/3}$. The gradient and the Hessian of the likelihood $\ell_{p,i}$ do not have a closed form. Fortunately, they can be approximated numerically in most statistical software.

In the last stage, we estimate θ_0 by ML after replacing $\boldsymbol{\beta}_0$ in the density function of y_i with $\hat{\boldsymbol{\beta}}_n$. Let $q_{n,i}(\theta,\ \hat{\boldsymbol{\beta}}_n) = \log\left(c_i(G_{1,i}(\hat{\boldsymbol{\beta}}_{n,1}),\ \dots,\ G_{k_n,i}(\hat{\boldsymbol{\beta}}_{n,k_n}),\ \theta)\right)$, where $\log c_i(u_1,\ \dots,\ u_{k_n}, \theta) = \sum_{p=1}^{k_n-1}\log(p\theta + 1) - (\theta+1)\sum_{p=1}^{k_n}\log u_p - (k_n + \frac{1}{\theta})\log(\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1)$. The objective function is $Q_n(\theta,\ \hat{\boldsymbol{\beta}}_n) = \frac{1}{n}\sum_{i=1}^n q_{n,i}(\theta,\ \hat{\boldsymbol{\beta}}_n)$. To compute $\partial_\theta q_{n,i}(\theta_0,\ \hat{\boldsymbol{\beta}}_n)$ and $\frac{\partial^2}{\partial\theta^2}q_{n,i}(\theta_0,\ \hat{\boldsymbol{\beta}}_n)$, we need the first and second derivatives of $\log c_i(u_1,\ \dots,\ u_{k_n}, \theta)$ that can be expressed as follows:

$$\partial_\theta \log c_i(u_1,\ \dots,\ u_{k_n}, \theta) = \sum_{p=1}^{k_n-1}\frac{p}{p\theta+1} - \sum_{p=1}^{k_n}\log u_p + \frac{\log\left(\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1\right)}{\theta^2} + \frac{(k_n+\frac{1}{\theta})\sum_{p=1}^{k_n} u_p^{-\theta}\log u_p}{\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1}$$

$$\frac{\partial^2}{\partial\theta^2}\log c_i(u_1,\ \dots,\ u_{k_n}, \theta) = \frac{(k_n+\frac{1}{\theta})\left(\sum_{p=1}^{k_n} u_p^{-\theta}\log u_p\right)^2}{\left(\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1\right)^2} - \frac{(k_n+\frac{1}{\theta})\sum_{p=1}^{k_n} u_p^{-\theta}(\log u_p)^2}{\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1} - \sum_{p=1}^{k_n}\left(\frac{p}{p\theta+1}\right)^2 - \frac{2\sum_{p=1}^{k_n} u_p^{-\theta}\log u_p}{\theta^2\left(\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1\right)} - \frac{2\log\left(\sum_{p=1}^{k_n} u_p^{-\theta} - k_n + 1\right)}{\theta^3}.$$

We define the following expressions: $\hat{A}_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta}q_{n,i}(\hat{\theta}_n,\ \hat{\boldsymbol{\beta}}_n)$,
$\hat{V}_n = \mathbb{V}_{\text{HAC}}(\frac{1}{\sqrt{n}}\sum_{i=1}^n \partial_\theta q_{n,i}(\hat{\theta}_n, \hat{\boldsymbol{\beta}}_n))$, and $\hat{\mathcal{E}}_{n,s} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \partial_\theta q_{n,i}(\hat{\theta}_n, \hat{\boldsymbol{\beta}}_n^{(s)})$, where $\hat{\boldsymbol{\beta}}_n^{(1)},\ \dots,\ \hat{\boldsymbol{\beta}}_n^{(\kappa)} \overset{i.i.d}{\sim} N(\hat{\boldsymbol{\beta}}_n,\ \hat{\mathbb{V}}(\hat{\boldsymbol{\beta}}_n))$. Let $\hat{\psi}_{n,s} = \frac{\sqrt{\hat{V}_n}\zeta_s + \hat{\mathcal{E}}_{n,s}}{\hat{A}_n}$, where $\zeta_1,\ \dots,\ \zeta_\kappa$ are independent variables from $N(0,1)$. The asymptotic CDF of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can be estimated by the empirical CDF of the sample: $\{\hat{\psi}_{n,s},\ s = 1,\ \dots,\ \kappa\}$. The estimator of the asymptotic variance of $\hat{\theta}_n$ is

$$\hat{\mathbb{V}}(\hat{\theta}_n) = \frac{1}{n\hat{A}_n^2}(\hat{V}_n + \frac{1}{\kappa-1}\sum_{s=1}^\kappa(\hat{\mathcal{E}}_{n,s} - \hat{\mathbb{E}}(\mathcal{E}_n))^2),$$

where $\hat{\mathbb{E}}(\mathcal{E}_n) = \frac{1}{\kappa}\sum_{s=1}^\kappa \hat{\mathcal{E}}_{n,s}$. The debiased estimator is given by

$$\theta_{n,\kappa}^* = \hat{\theta}_n - \hat{\mathbb{E}}(\mathcal{E}_n)/(\sqrt{n}\hat{A}_n).$$

Let $\hat{\psi}_{n,s}^* = \frac{\sqrt{\hat{V}_n^*}\zeta_s + \hat{\mathcal{E}}_{n,s}^* - \hat{\mathbb{E}}(\mathcal{E}_n^*)}{\hat{A}_n^*}$, where $\hat{A}_n^*$, $\hat{V}_n^*$, $\hat{\mathcal{E}}_{n,s}^*$, and $\hat{\mathbb{E}}(\mathcal{E}_n^*)$ are defined as $\hat{A}_n$, $\hat{V}_n$, $\hat{\mathcal{E}}_{n,s}$, and $\hat{\mathbb{E}}(\mathcal{E}_n)$, respectively, with the difference that they are computed using $\theta_{n,\kappa}^*$ and not $\hat{\theta}_n$. We can estimate the asymptotic CDF of $\sqrt{n}(\theta_{n,\kappa}^* - \theta_0)$ by the empirical CDF of the sample: $\{\hat{\psi}_{n,s},\ s = 1,\ \dots,\ \kappa\}$.

S.2.2 Estimates of the Asymptotic Distribution of the Debiased Estimators

This section presents the estimates of the asymptotic CDF of $\Delta_{n,\kappa}^* := \sqrt{n}(\boldsymbol{\theta}_{n,\kappa}^* - \theta_0)$, where $\boldsymbol{\theta}_{n,\kappa}^*$ is the debiased estimator. In contrast to the case of the classical plug-in estimator, the true sampling CDFs are asymptotically centered at zero because $\mathbb{E}(\Delta_{n,\kappa}^*)$ converges to zero asymptotically. Overall, the results demonstrate that the estimator of the CDF of $\Delta_{n,\kappa}^*$, as outlined in Theorem 4.3, performs well.

[3]See ANDREWS, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica: Journal of the Econometric Society*, 59, 817–858.

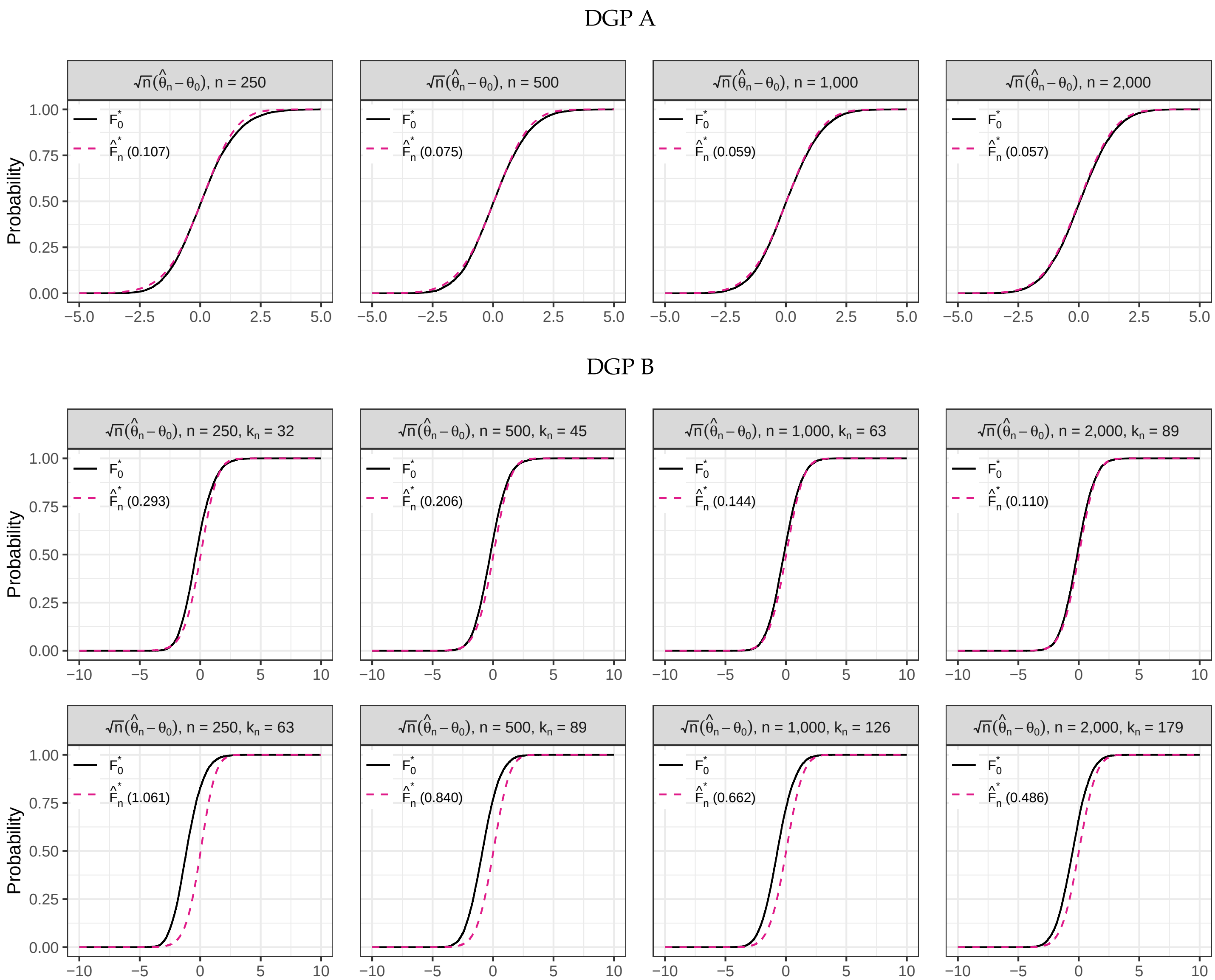
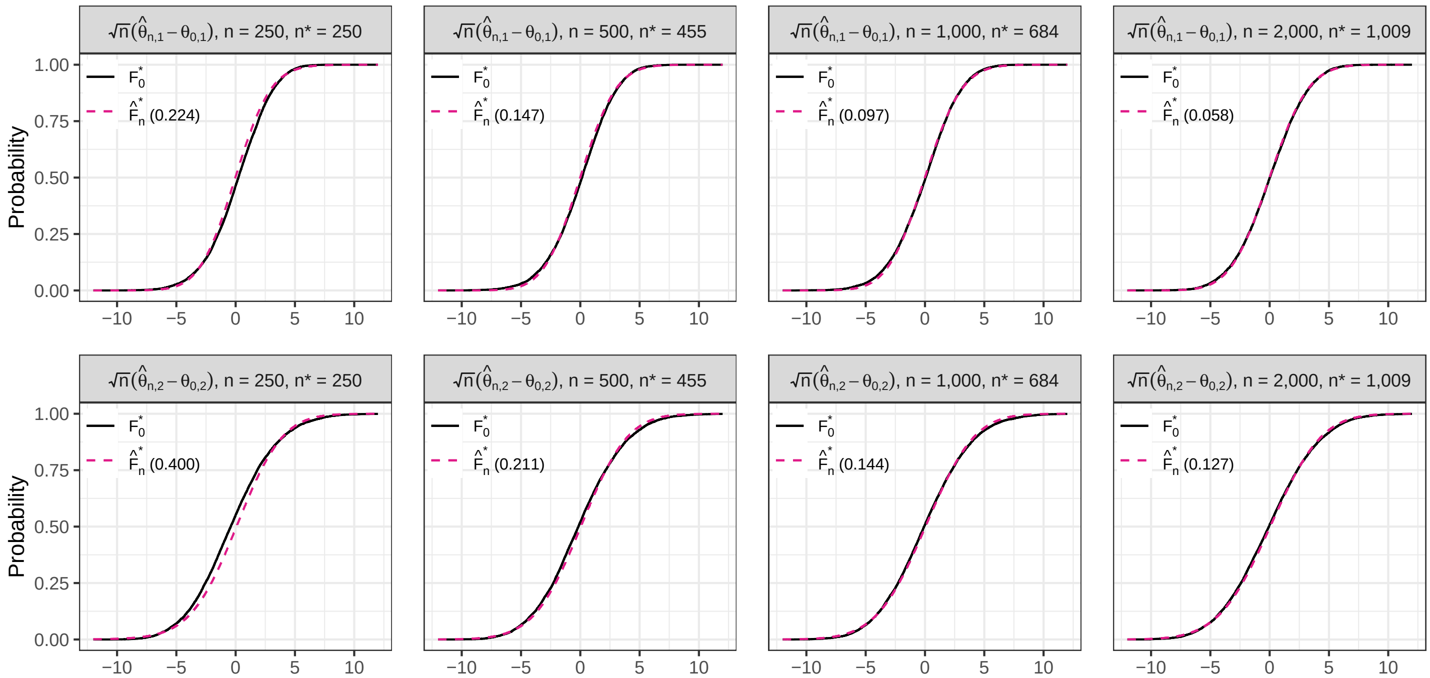


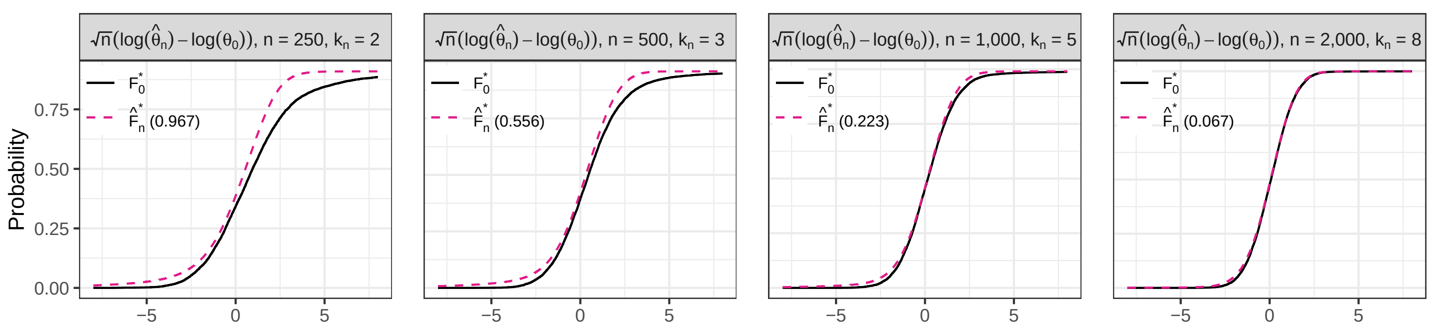
Figure S.5: Monte Carlo Simulations: Estimates of asymptotic CDFs (DGPs A and B)

This figure displays average estimates of the asymptotic CDF of $\sqrt{n}(\boldsymbol{\theta}^*_{n,\kappa} - \theta_0)$ for DGPs A and B. F_0^* represents the true sampling CDF whereas $\hat{F}_n^*$ corresponds to the average estimate of the CDF using our simulation approach. The L_1-Wasserstein distance between each estimated CDF and F_0^* is enclosed in parentheses.

DGP C



DGP D: Mean Correction



DGP D: Median Correction

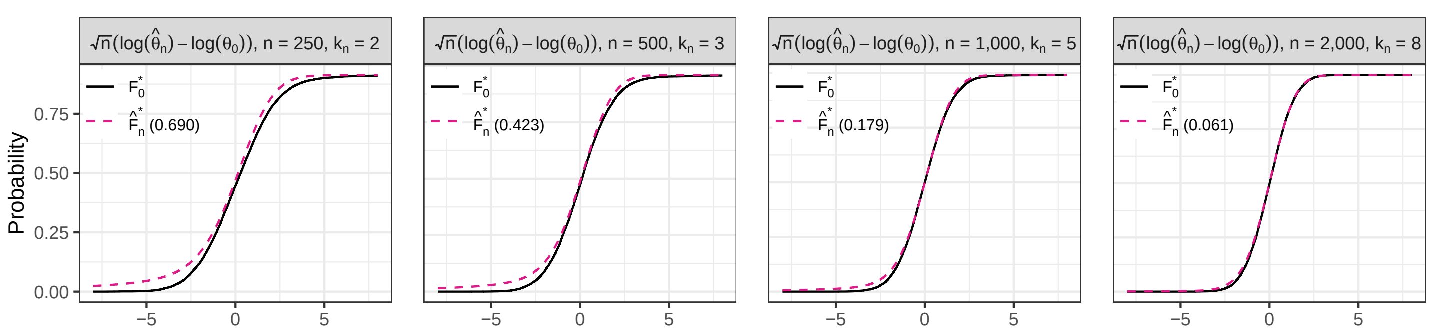


Figure S.6: Monte Carlo Simulations: Estimates of asymptotic CDFs (DGPs C and D)

This figure displays average estimates of the asymptotic CDF of $\sqrt{n}(\boldsymbol{\theta}^*_{n,\kappa} - \theta_0)$ for DGPs A and B. F_0^* represents the true sampling CDF whereas $\hat{F}_n^*$ corresponds to the average estimate of the CDF using our simulation approach. The L_1-Wasserstein distance between each estimated CDF and F_0^* is enclosed in parentheses.

S.3 Supplementary Materials on the Application

S.3.1 Data Summary

Our dependent variable is the weekly fast-food consumption frequency, measured by the reported frequency (in days) of fast-food restaurant visits in the past week. We control for 25 observable characteristics in $\mathbf{X}_r$ such as students' gender, grade, race, weekly allowance, and parent's education and occupation. On average, students report consuming fast food 2.35 days per week. (see Table S.1).

Table S.1: Data Summary

Statistic	Mean	St. Dev.	Min	Max
Fast food consumption	2.353	1.762	0	7
Female	0.501	0.500	0	1
Age	16.628	1.554	12	21
Hispanic	0.200	0.400	0	1
Grade 7–8	0.100	0.300	0	1
Grade 9–10	0.230	0.421	0	1
Grade 11–12	0.533	0.499	0	1
Race (White)				
Black	0.142	0.349	0	1
Asian	0.138	0.345	0	1
Other	0.117	0.321	0	1
With parents	0.675	0.468	0	1
Allowance per week	7.893	11.609	0	95
Mother Education (High school)				
$<$ High school	0.146	0.353	0	1
$>$ High school and not graduated	0.173	0.378	0	1
$>$ High school and graduated	0.226	0.418	0	1
Missing	0.124	0.330	0	1
Father Education (High school)				
$<$ High school	0.124	0.329	0	1
$>$ High school and not graduated	0.137	0.344	0	1
$>$ High school and graduated	0.202	0.402	0	1
Missing	0.284	0.451	0	1
Mother Job (None)				
Professional	0.157	0.364	0	1
Other	0.623	0.485	0	1
Missing	0.088	0.283	0	1
Father Job (None)				
Professional	0.053	0.223	0	1
Other	0.663	0.473	0	1
Missing	0.240	0.427	0	1

This table presents the mean, standard deviation (St. Dev.), minimum, and maximum of the variables used in the empirical application. For the categorical explanatory variables, the level in parentheses is set as the reference level. "With parents" is a dummy variable taking 1 if the student lives with their mother and father.

S.3.2 Estimation and Inference

The following table displays the full results of estimations of the peer effect model.

Table S.2: Estimation Results: OLS approach

Fixed Effects	No		Yes	
	Coef	Sd. Err	Coef	Sd. Err
Peer effects: $\theta_{0,1}$	0.192	0.031	0.150	0.032
Individual characteristics: $\boldsymbol{\theta}_{0,2}$				
Female	−0.158	0.075	−0.161	0.074
Age	0.109	0.040	0.083	0.040
Hispanic	0.289	0.130	0.103	0.147
Grade 7–8	−0.033	0.257	0.102	0.260
Grade 9–10	0.020	0.168	0.020	0.169
Grade 11–12	0.258	0.116	0.205	0.117
Race (White)				
Black	0.075	0.145	−0.085	0.167
Asian	0.302	0.144	0.134	0.167
Other	−0.118	0.138	−0.145	0.139
With parents	−0.001	0.134	0.002	0.134
Allowance per week	0.008	0.003	0.006	0.003
Mother Education (High school)				
< High school	0.101	0.119	0.084	0.119
> High school and non graduated	0.054	0.100	0.023	0.100
> High school and graduated	0.132	0.106	0.086	0.106
Missing	−0.035	0.191	−0.103	0.190
Father Education (High school)				
< High school	−0.251	0.128	−0.244	0.127
> High school and non graduated	−0.100	0.112	−0.112	0.112
> High school and graduated	−0.005	0.109	−0.038	0.110
Missing	0.016	0.186	−0.024	0.186
Mother Job (None)				
Professional	0.015	0.133	−0.007	0.133
Other	0.065	0.101	0.061	0.101
Missing	0.376	0.227	0.441	0.226
Father Job (None)				
Professional	−0.247	0.224	−0.281	0.223
Other	−0.230	0.165	−0.234	0.164
Missing	−0.252	0.250	−0.228	0.250
Contextual peer effects: $\boldsymbol{\theta}_{0,3}$				
Female	0.044	0.120	−0.001	0.121
Age	−0.030	0.022	−0.008	0.023
Hispanic	−0.091	0.195	−0.189	0.203
Grade 7–8	−0.268	0.274	−0.281	0.277
Grade 9–10	−0.189	0.209	−0.107	0.209
Grade 11–12	−0.028	0.189	0.028	0.188
Race (White)				
Black	0.165	0.198	0.129	0.206
Asian	−0.118	0.187	−0.122	0.196
Other	−0.511	0.228	−0.470	0.228
With parents	−0.199	0.216	−0.230	0.218
Allowance per week	0.006	0.005	0.003	0.005
Mother Education (High school)				
< High school	0.417	0.196	0.359	0.196
> High school and non graduated	−0.163	0.170	−0.226	0.171
> High school and graduated	−0.065	0.178	−0.189	0.180
Missing	−0.288	0.373	−0.386	0.372
Father Education (High school)				
< High school	−0.090	0.216	−0.111	0.217
> High school and non graduated	0.140	0.176	0.084	0.179
> High school and graduated	0.093	0.177	0.052	0.182
Missing	0.289	0.324	0.278	0.324
Mother Job (None)				
Professional	−0.217	0.223	−0.246	0.224
Other	−0.249	0.171	−0.272	0.172
Missing	−0.193	0.427	−0.105	0.428
Father Job (None)				
Professional	0.484	0.370	0.111	0.376
Other	0.368	0.268	0.238	0.274
Missing	0.044	0.424	−0.116	0.430

For the categorical variables, the level in parentheses is set as the reference level.

Table S.3: Estimation Results: Classical and Optimal GMM approaches

Model	CIV		CIV		OIV		OIV	
Fixed Effects	No		Yes		No		Yes	
	Coef	Sd. Err	Coef	Sd. Err	Coef	Sd. Err	Coef	Sd. Err
Peer effects: $\theta_{0,1}$	0.149	0.160	0.081	0.169	−0.065	0.287	0.016	0.208
Individual characteristics: $\boldsymbol{\theta}_{0,2}$								
Female	−0.156	0.075	−0.158	0.075	−0.145	0.077	−0.154	0.075
Age	0.111	0.040	0.085	0.040	0.121	0.042	0.087	0.041
Hispanic	0.288	0.130	0.095	0.148	0.284	0.132	0.088	0.149
Grade 7–8	−0.033	0.258	0.105	0.260	−0.035	0.261	0.108	0.261
Grade 9–10	0.019	0.168	0.018	0.169	0.017	0.171	0.016	0.169
Grade 11–12	0.258	0.116	0.203	0.117	0.261	0.118	0.201	0.117
Race (White)								
Black	0.073	0.145	−0.091	0.168	0.066	0.147	−0.097	0.168
Asian	0.305	0.144	0.133	0.167	0.318	0.147	0.132	0.167
Other	−0.123	0.139	−0.153	0.141	−0.144	0.142	−0.160	0.142
With parents	−0.008	0.136	−0.006	0.135	−0.042	0.143	−0.015	0.137
Allowance per week	0.008	0.003	0.006	0.003	0.008	0.003	0.006	0.003
Mother Education (High school)								
< High school	0.102	0.120	0.085	0.119	0.110	0.121	0.087	0.119
> High school and not graduated	0.057	0.100	0.025	0.100	0.070	0.103	0.026	0.101
> High school and graduated	0.131	0.106	0.082	0.107	0.128	0.108	0.079	0.107
Missing	−0.045	0.194	−0.121	0.195	−0.094	0.204	−0.138	0.198
Father Education (High school)								
< High school	−0.252	0.128	−0.246	0.127	−0.260	0.130	−0.247	0.128
> High school and not graduated	−0.100	0.112	−0.113	0.112	−0.101	0.113	−0.114	0.112
> High school and graduated	−0.006	0.109	−0.040	0.111	−0.011	0.111	−0.042	0.111
Missing	0.019	0.187	−0.020	0.186	0.037	0.190	−0.017	0.186
Mother Job (None)								
Professional	0.011	0.134	−0.014	0.134	−0.009	0.137	−0.021	0.135
Other	0.066	0.101	0.061	0.101	0.069	0.103	0.061	0.102
Missing	0.381	0.228	0.452	0.228	0.404	0.232	0.463	0.229
Father Job (None)								
Professional	−0.242	0.224	−0.278	0.224	−0.219	0.228	−0.274	0.224
Other	−0.225	0.165	−0.228	0.165	−0.203	0.169	−0.222	0.166
Missing	−0.255	0.251	−0.230	0.250	−0.267	0.254	−0.231	0.251
Contextual peer effects: $\boldsymbol{\theta}_{0,3}$								
Female	0.037	0.123	−0.013	0.124	0.003	0.130	−0.024	0.126
Age	−0.025	0.029	0.001	0.031	0.000	0.040	0.009	0.034
Hispanic	−0.075	0.203	−0.169	0.209	0.002	0.222	−0.150	0.213
Grade 7–8	−0.264	0.274	−0.273	0.278	−0.246	0.278	−0.266	0.279
Grade 9–10	−0.186	0.209	−0.099	0.210	−0.174	0.212	−0.092	0.211
Grade 11–12	−0.011	0.199	0.056	0.201	0.073	0.222	0.083	0.207
Race (White)								
Black	0.172	0.199	0.141	0.208	0.205	0.205	0.153	0.209
Asian	−0.106	0.192	−0.101	0.202	−0.047	0.205	−0.082	0.206
Other	−0.519	0.230	−0.483	0.230	−0.560	0.237	−0.495	0.232
With parents	−0.212	0.221	−0.247	0.222	−0.276	0.235	−0.264	0.224
Allowance per week	0.007	0.006	0.004	0.005	0.009	0.006	0.004	0.006
Mother Education (High school)								
< High school	0.420	0.197	0.359	0.196	0.438	0.200	0.359	0.197
> High school and not graduated	−0.159	0.170	−0.223	0.172	−0.139	0.174	−0.220	0.172
> High school and graduated	−0.053	0.184	−0.178	0.182	0.010	0.199	−0.168	0.184
Missing	−0.264	0.383	−0.351	0.382	−0.146	0.409	−0.319	0.387
Father Education (High school)								
< High school	−0.112	0.231	−0.144	0.231	−0.223	0.264	−0.175	0.239
> High school and not graduated	0.124	0.186	0.054	0.193	0.043	0.209	0.025	0.201
> High school and graduated	0.082	0.181	0.035	0.186	0.029	0.192	0.020	0.189
Missing	0.282	0.325	0.265	0.326	0.251	0.331	0.254	0.327
Mother Job (None)								
Professional	−0.225	0.225	−0.262	0.227	−0.265	0.232	−0.276	0.229
Other	−0.245	0.172	−0.268	0.173	−0.227	0.175	−0.265	0.173
Missing	−0.209	0.431	−0.125	0.431	−0.286	0.445	−0.144	0.433
Father Job (None)								
Professional	0.501	0.375	0.119	0.377	0.587	0.392	0.125	0.378
Other	0.382	0.273	0.256	0.277	0.452	0.287	0.273	0.280
Missing	0.061	0.429	−0.091	0.434	0.146	0.444	−0.067	0.438

For the categorical variables, the level in parentheses is set as the reference level.

Table S.4: Estimation Results: Many Instrument Approaches

Model	IV-MI		IV-MI		DIV-MI		DIV-MI	
Fixed Effects	No		Yes		No		Yes	
	Coef	Sd. Err	Coef	Sd. Err	Coef	Sd. Err	Coef	Sd. Err
Peer effects: $\theta_{0,1}$	0.276	0.063	0.208	0.067	0.300	0.063	0.218	0.067
Individual characteristics: $\boldsymbol{\theta}_{0,2}$								
Female	−0.162	0.072	−0.164	0.070	−0.163	0.072	−0.165	0.070
Age	0.105	0.038	0.082	0.038	0.104	0.038	0.082	0.038
Hispanic	0.290	0.126	0.109	0.135	0.290	0.126	0.110	0.135
Grade 7–8	−0.032	0.252	0.100	0.241	−0.030	0.252	0.102	0.241
Grade 9–10	0.021	0.157	0.021	0.154	0.020	0.157	0.023	0.154
Grade 11–12	0.257	0.118	0.207	0.119	0.255	0.118	0.209	0.119
Race (White)								
Black	0.078	0.147	−0.080	0.158	0.080	0.147	−0.078	0.158
Asian	0.297	0.134	0.135	0.151	0.296	0.134	0.136	0.151
Other	−0.110	0.129	−0.138	0.128	−0.108	0.129	−0.135	0.128
With parents	0.012	0.128	0.010	0.127	0.016	0.129	0.011	0.128
Allowance per week	0.008	0.003	0.006	0.003	0.008	0.003	0.006	0.003
Mother Education (High school)								
< High school	0.098	0.113	0.082	0.114	0.097	0.114	0.082	0.114
> High school and not graduated	0.048	0.093	0.022	0.093	0.047	0.094	0.021	0.093
> High school and graduated	0.133	0.095	0.089	0.094	0.134	0.095	0.089	0.094
Missing	−0.015	0.179	−0.088	0.178	−0.008	0.179	−0.085	0.178
Father Education (High school)								
< High school	−0.248	0.114	−0.243	0.112	−0.247	0.114	−0.244	0.112
> High school and not graduated	−0.100	0.102	−0.111	0.102	−0.100	0.102	−0.111	0.102
> High school and graduated	−0.003	0.103	−0.036	0.105	−0.004	0.103	−0.035	0.105
Missing	0.009	0.180	−0.027	0.179	0.007	0.180	−0.025	0.179
Mother Job (None)								
Professional	0.023	0.122	−0.001	0.124	0.024	0.122	0.002	0.124
Other	0.064	0.093	0.060	0.095	0.063	0.093	0.062	0.095
Missing	0.366	0.216	0.432	0.213	0.362	0.216	0.429	0.213
Father Job (None)								
Professional	−0.256	0.192	−0.284	0.196	−0.261	0.192	−0.286	0.196
Other	−0.238	0.136	−0.239	0.139	−0.241	0.137	−0.239	0.139
Missing	−0.248	0.230	−0.227	0.231	−0.246	0.230	−0.226	0.231
Contextual peer effects: $\boldsymbol{\theta}_{0,3}$								
Female	0.058	0.111	0.010	0.112	0.062	0.111	0.013	0.112
Age	−0.040	0.020	−0.015	0.021	−0.043	0.020	−0.016	0.021
Hispanic	−0.121	0.167	−0.206	0.180	−0.129	0.167	−0.210	0.180
Grade 7–8	−0.275	0.241	−0.287	0.241	−0.279	0.241	−0.286	0.242
Grade 9–10	−0.194	0.179	−0.114	0.180	−0.193	0.179	−0.112	0.180
Grade 11–12	−0.061	0.171	0.003	0.173	−0.067	0.172	0.003	0.173
Race (White)								
Black	0.152	0.190	0.119	0.197	0.147	0.190	0.113	0.197
Asian	−0.141	0.173	−0.139	0.181	−0.149	0.173	−0.145	0.181
Other	−0.495	0.177	−0.459	0.180	−0.489	0.177	−0.456	0.180
With parents	−0.174	0.194	−0.215	0.198	−0.165	0.194	−0.215	0.198
Allowance per week	0.005	0.005	0.003	0.005	0.005	0.005	0.003	0.005
Mother Education (High school)								
< High school	0.410	0.180	0.359	0.177	0.409	0.180	0.361	0.177
> High school and not graduated	−0.171	0.147	−0.229	0.146	−0.173	0.147	−0.228	0.146
> High school and graduated	−0.090	0.161	−0.199	0.166	−0.096	0.162	−0.199	0.167
Missing	−0.335	0.351	−0.415	0.354	−0.345	0.352	−0.419	0.354
Father Education (High school)								
< High school	−0.046	0.182	−0.083	0.180	−0.031	0.183	−0.082	0.180
> High school and not graduated	0.172	0.155	0.110	0.158	0.184	0.156	0.115	0.158
> High school and graduated	0.113	0.157	0.065	0.163	0.122	0.158	0.065	0.163
Missing	0.301	0.241	0.288	0.243	0.308	0.241	0.289	0.242
Mother Job (None)								
Professional	−0.202	0.200	−0.233	0.201	−0.196	0.200	−0.231	0.201
Other	−0.256	0.150	−0.276	0.153	−0.254	0.150	−0.275	0.153
Missing	−0.163	0.385	−0.088	0.384	−0.150	0.386	−0.082	0.384
Father Job (None)								
Professional	0.450	0.289	0.105	0.298	0.438	0.289	0.106	0.298
Other	0.341	0.187	0.223	0.199	0.332	0.187	0.219	0.198
Missing	0.010	0.322	−0.137	0.332	0.000	0.322	−0.142	0.331

For the categorical variables, the level in parentheses is set as the reference level.