SUPPLEMENT TO "IDENTIFYING SOCIAL INTERACTIONS THROUGH CONDITIONAL VARIANCE RESTRICTIONS: ADDITIONAL EMPIRICAL RESULTS, SPECIFICATION TESTS, AND POWER CALCULATIONS

BY BRYAN S. GRAHAM

This document provides some supplementary empirical and theoretical results for the paper titled "Identifying Social Interactions Through Conditional Variance Restrictions." All notation, unless explicitly stated otherwise, is as defined in the paper, and numbering of text elements continues in sequence with the paper.

A. SPECIFICATION TESTING

SINCE THE ESTIMATOR PROPOSED in the paper is based on a simple conditional moment restriction, standard approaches to specification testing are available (e.g., the Sargan–Hansen test of overidentifying restrictions). However, as in any application, it is helpful to think about specific directions of misspecification and to construct tests accordingly.

As an example, note that an implication of Assumption 1.2 (as stated in the paper) is that class size/type and teacher characteristics are 'stochastically separable' in the production of academic achievement. This appendix outlines a test for this assumption.

Assume that teachers have L latent attributes $\underline{a}_c = (a_{c1}, \ldots, a_{cL})'$. Conditional random assignment ensures that

$$\mathbb{V}(\underline{a}_c | W_{1c}, W_{2c}) = \mathbb{V}(\underline{a}_c | W_{1c}).$$

The relative importance of each attribute for realized *teaching effectiveness*, however, may vary with class type, for example,

$$(15) \qquad \alpha_c = (\kappa_0 + \underline{a}_c' \lambda_1) \cdot W_{2c} + \underline{a}_c' \lambda_0 \cdot (1 - W_{2c}), \quad \lambda_0 \neq \lambda_1.$$

In the notation of the main paper $A_c(1) = \kappa_0 + \underline{a}_c' \lambda_1$ and $A_c(0) = \underline{a}_c' \lambda_0$. In this particular model of teacher effectiveness, Assumption 1.2 requires that $\lambda_1 = \lambda_0$. More generally we could allow the factor loadings, λ_1 and λ_0, to be teacher-specific; however, this would complicate the analysis which follows (in that case, Assumption 1.2 would follow if $\lambda_{1c} \sim \lambda_{0c}$).

An implication of (15) is that even under random assignment of teachers to class type, the conditional variance of teacher effectiveness will differ across small and regular classrooms (even though the distribution of underlying latent teacher attributes will not). Formally,

$$\mathbb{V}(\alpha_c | W_{1c}, W_{2c} = 1) - \mathbb{V}(\alpha_c | W_{1c}, W_{2c} = 0)$$
$$= \lambda_1' \mathbb{V}(\underline{a}_c | W_{1c}) \lambda_1 - \lambda_0' \mathbb{V}(\underline{a}_c | W_{1c}) \lambda_0 \neq 0 \qquad \text{(for } \lambda_1 \neq \lambda_0\text{)},$$

which violates Assumption 1.2 of the paper. Assessing the plausibility of Assumption 1.2 therefore requires both consideration of the assignment process as well as the nature of the educational production function.[30]

Relatively little is known about the educational production process. One view suggests that class size and some underlying notion of teacher ability are complementary. This would imply that

$$\frac{\sigma_{A(1)}^2}{\sigma_{A(0)}^2} = \frac{\lambda_1' \mathbb{V}(\underline{a}_c | W_{1c}) \lambda_1}{\lambda_0' \mathbb{V}(\underline{a}_c | W_{1c}) \lambda_0} > 1$$

or that holding the distribution of teacher characteristics fixed, teacher effectiveness is more variable in large relative to small classrooms. With complementarity we would expect that moving a fixed population of teachers to larger classrooms would, in addition to reducing *average* teacher effectiveness, increase its *variance.* In this case all teachers would perform relatively similarly in small classrooms with differences in teacher effectiveness only emerging in larger classrooms. Alternatively, teacher ability and class size could be substitutable, with individual teacher characteristics being unimportant in large classrooms, because, for example, anyone can effectively execute "chalk and talk."

A simple and direct test for substitutability/complementarity bias its to compare an estimate of γ^2 based on a random sample from a subpopulation of groups with large amounts of heterogeneity in an observed teacher attribute with an estimate based on a random sample from a subpopulation of groups with little heterogeneity. If size and the teacher attribute are complementary, then the estimate based on the first subpopulation should be smaller than that based on the second. If the teacher attribute and class size are substitutes, the opposite pattern will occur.

In the Project STAR data set the only observed teacher covariate that is significantly related to test scores is years of teaching experience. I divide Project STAR schools (and hence classrooms) into two sets: in the first set, the standard deviation of years teaching experience is greater than or equal to 5; in the second set, it is less than 5. This partition is used to form subsamples with high and low degrees of heterogeneity in teacher quality.

Table II reports separate estimates of γ^2 using these two subsamples. The discussion emphasizes the math achievement results since those for reading achievement are not well identified, with first stage F-statistics all below 10. Column 1 reports the GMM estimate of γ^2 based on a comparison across small and large classrooms in schools with "lots" of heterogeneity in years of teaching experience. Column 2 reports the estimate based on classrooms in schools with little experience heterogeneity. The two estimates are precisely estimated and similar in magnitude, consistent with the null of separability.

---

[30]This is, of course, no different than in other areas of applied economics where assumptions on technology are often crucial for achieving identification.

TABLE II

GMM Estimates of γ^2 Based on Excess Variance Contrasts Across High and Low Experience Heterogeneity Subsamples[a]

	(1) High Heterogeneity (HH) $\text{Std}(\text{Exp}_c) \geq 5$	(2) Low Heterogeneity (LH) $\text{Std}(\text{Exp}_c) < 5$	(3) Combined Sample
Panel A: Math Achievement			
γ^2	3.5631	3.3478	3.4457
	(1.6247)	(1.2637)	(1.0136)
Regular-with-aide	−0.0176	0.0173	0.0186
	(0.0361)	(0.0238)	(0.0233)
High heterogeneity ×			−0.0372
regular-with-aide	—	—	(0.0444)
$F_{(df1, df2)}$ 1st Stage	$2.94_{(1,104)}$	$49.42_{(1,130)}$	$31.27_{(2,234)}$
p-value $H_0: \gamma^2_{\text{HH}} = \gamma^2_{\text{LH}}$	—	—	$p = 0.9037$
Panel B: Reading Achievement			
γ^2	9.5279	2.1262	5.1881
	(5.7685)	(1.6931)	(2.4785)
Regular-with-aide	0.0428	0.0076	0.0562
	(0.0633)	(0.0329)	(0.0461)
High heterogeneity ×			−0.0528
regular-with-aide	—	—	(0.0776)
$F_{(df1, df2)}$ 1st Stage	$5.08_{(1,104)}$	$6.14_{(1,130)}$	$5.61_{(2,234)}$
p-value $H_0: \gamma^2_{\text{HH}} = \gamma^2_{\text{LH}}$	—	—	$p = 0.1214$
Number of classrooms	140	177	317
School fixed effects?	Yes	Yes	Yes

[a]Columns 1 and 2 report GMM estimates of γ^2 based on (14) and subsamples exhibiting high and low degrees of heterogeneity in years of teaching experience, respectively. The null that $\gamma^2_{\text{LH}} = \gamma^2_{\text{HH}}$ is tested using the Sargan–Hansen test of overidentifying restrictions associated with the column 3 estimates, where a binary variable for whether a classroom is of the small type and its interaction with a binary variable for being in a high heterogeneity school serve as excluded instruments.

Column 3 reports two-step GMM estimates of γ^2 using the entire sample with the small class type dummy and its interaction with a dummy for belonging to the high heterogeneity subsample ($\text{Std}(\text{Exp}_c) \geq 5$ years) serving as excluded instruments. Row 3 reports the p-value for a Sargan–Hansen test of the null that the high heterogeneity and low heterogeneity estimates of γ^2 are equal. There is little evidence of quantitatively important bias due to nonseparability of teacher characteristics and class size in the educational production function.

This conclusion is consistent with the finding that teacher instructional practices are not sensitive to modest variations in class size (Betts and Shkolnik (1999), Rice (1999)). Evidence from direct observation of STAR classrooms also found "that teacher practices did not change substantially regardless of class type assignment" (Evertson and Randolph (1989, p. 102); cf. Word et al.

(1990)). The combination of quantitative and qualitative evidence suggest that the separability null is a reasonable.

An important dividend associated with the method-of-moments representation of the estimation procedure is that specification testing is standard and can be guided by the economics of the application under consideration.

B. SENSITIVITY ANALYSIS

Let $\mathbb{E}[G_c^w|W_{1c}, W_{2c}] = W_{1c}'\vartheta_1 + W_{2c}'\vartheta_2$ denote the first stage population regression of G_c^w onto W_{1c} and W_{2c}. Using Equation (14) in the paper we can show that the probability limit of $\widehat{\gamma}^2$ is given by

$$(16) \qquad \gamma_*^2 = \gamma_0^2 + \frac{\sigma_{A(0)}^2(\sigma_{A(1)}^2/\sigma_{A(0)}^2 - 1)}{\vartheta_2}.$$

Assume that there are no social interactions ($\gamma_0^2 = 1$). Then (16) can be combined with assumptions on $\sigma_{A(0)}^2$ to back out the degree of nonseparability between teacher attributes and class type that would be required to produce (large sample) estimates of γ_*^2 of the size reported in Table I.

Table III reports the results of exercises of this type. To calibrate the experiments, note that $\sigma_{A(0)}$ equals the change in test scores associated with a 1 standard deviation change in teacher effectiveness *in regular and regular-with-aide classrooms.* The relevant distribution is the *within-school* distribution of teacher effectiveness, since the between-school variation in test scores has already been purged from the data. Rockoff (2004), using panel data methods, simple deconvolution procedures to deal with measurement error, and a sample of normal sized classrooms from New Jersey, estimated $\sigma_{A(0)}$ to be about 0.1. The lower bound estimates of Rivkin, Hanusheck, and Kain (2005, Table III, column 3, p. 434) are similar to those of Rockoff (2004). Aaronson, Barrow, and Sander's (2003) research, using Chicago Public Schools data, sug-

TABLE III

ROBUSTNESS TO VIOLATIONS OF STOCHASTIC SEPARABILITY

	$\sigma_{A(0)}$	$\frac{\sigma_{A(1)}}{\sigma_{A(0)}} = \sqrt{1 + \frac{\vartheta_2}{\sigma_{A(0)}^2}(\gamma_*^2 - \gamma_0^2)}$
Panel A:	0.10	2.59
	0.20	1.56
	0.30	1.28
	0.40	1.16
Panel B:		
ϑ_2	0.0231	
γ_*^2	3.4691	

gests a somewhat higher value for $\sigma_{A(0)}$. A conservative upper bound for $\sigma_{A(0)}$ based on existing evidence is about 0.3.

For $\sigma_{A(0)} = 0.1$, the typical difference in effectiveness across a pair of teachers would have to be roughly 2.5 times larger in small versus large classrooms to produce γ_*^2 estimates of the size reported in Table I, if in fact there were no peer effects. This is an implausibly large number. For $\sigma_{A(0)} = 0.3$, the difference would have to be 1.3 times larger, still quite a large effect. Overall identification appears to be strong enough to ensure a reasonable amount of robustness to violations of Assumption 1.2.

C. COMPARISON OF THE EXCESS VARIANCE TEST WITH CONVENTIONAL REGRESSION-BASED TESTS

The most common and arguably current best practice test for social interactions is a reduced form test for excess sensitivity (e.g., Sacerdote (2001), Angrist and Lang (2004)). This method exploits random assignment, or conditional random assignment, of individuals to groups to motivate simple least squares-based tests for social interactions. These tests are attractive since their plausibility is straightforward to evaluate and they are easy to implement. Graham and Hahn (2005) provided a formal overview of this approach. This appendix outlines the intuition behind such tests, applies them to the Project STAR data set, and compares them with the excess variance test developed in the main paper.

Implementing these tests requires that in addition to outcomes, we observe a $K \times 1$ vector of individual-level characteristics, R_{ci}. This allows the individual heterogeneity term to be decomposed into observable and unobservable components, $\varepsilon_{ci} = R'_{ci}\eta + \epsilon_{ci}$ with $\mathbb{E}^*[\epsilon_{ci}|R_{ci}] = 0$.[31]

Substituting $\varepsilon_{ci} = R'_{ci}\eta + \epsilon_{ci}$ into (1) and rearranging to partition achievement into its within- and between-group components yields the reduced form

$$(17) \qquad Y_{ci} = \overline{R}'_c \eta_0 \gamma_0 + (R_{ci} - \overline{R}_c)' \eta_0 + U_{ci}$$
$$= \overline{R}'_c \pi_{b0} + \widetilde{R}'_{ci} \pi_{w0} + U_{ci},$$

where $U_{ci} = \alpha_c + \gamma_0 \overline{\epsilon}_c + (\epsilon_{ci} - \overline{\epsilon}_c)$, $\overline{R}_c$ is the group mean of R_{ci}, and $\widetilde{R}_{ci} = R_{ci} - \overline{R}_c$.

Under random assignment of students to classrooms $\mathbb{E}[\alpha_c|\underline{R}_c] = \mathbb{E}[\alpha_c]$ and a least squares regression of Y_{ci} on a constant, $\overline{R}_c$ and $\widetilde{R}_{ci}$ identifies $\pi_0 = (\pi'_{b0}, \pi'_{w0})' = ((\gamma_0, 1) \otimes \eta'_0)'$. The null hypothesis of no social interactions can be assessed by testing the restriction $\pi_{b0} = \pi_{w0}$. Positive social interactions imply that $\pi_{b0} > \pi_{w0}$ or that there is *excess between-group sensitivity* in outcomes

---

[31]For simplicity, I assume that R_{ci} is mean zero.

to between-group variation in characteristics. Note that π_{b0} and π_{w0} are identical to the coefficients in the within- and between-group regressions of Y on R, and hence, formally, the excess sensitivity test is a Hausman and Taylor (1981) test, although its motivation and interpretation are quite different (Graham and Hahn (2005)).

Table IV implements the excess sensitivity test for social interactions using Project STAR math and reading test score data. The table reports least squares estimates of π_{b0}, π_{w0}, and their difference. Included in R_{ci} are dummies for gender, race, and eligibility for free/reduced price school lunch. Also included in the regression are class type and school dummies (coefficients not reported).

To facilitate comparisons, the between- and within-group coefficients are reported side-by-side in columns 1 and 2, with column 3 giving the difference.

TABLE IV

VARIABLE-BY-VARIABLE TESTS FOR EXCESS SENSITIVITY IN NORMALIZED KINDERGARTEN SAT MATH AND READING SCORES[a]

Excess Sensitivity Tests	(1) π_b	(2) π_w	(3) $\pi_b - \pi_w$
Panel A: Math			
π_{BLACK}	−0.2106	−0.3752	0.1646
	(0.4882)	(0.0531)	(0.4980)
π_{GIRL}	0.5274	0.1187	0.4087
	(0.1861)	(0.0231)	(0.1859)*
$\pi_{\text{FREELUNCH}}$	−0.5620	−0.4109	−0.1511
	(0.2026)	(0.0280)	(0.2044)
Omnibus Test Results	*F* Statistic		*p* Value
$F_{(df1,df2)}$ $(H_0 : \pi_b = \pi_w)$	$1.89_{(3,316)}$		0.1304
$F_{(df1,df2)}$ $(H_0 : \pi_b = 2 \cdot \pi_w)$	$1.55_{(3,316)}$		0.2027
Panel B: Reading			
π_{BLACK}	−0.7425	−0.2499	−0.426
	(0.4112)	(0.0540)	(0.4219)
π_{GIRL}	0.4955	0.1520	0.3435
	(0.1726)	(0.0250)	(0.1744)*
$\pi_{\text{FREELUNCH}}$	−0.5000	−0.4534	−0.0466
	(0.1821)	(0.0283)	(0.1819)
Omnibus Test Results	*F* Statistic		*p* Value
$F_{(df1,df2)}$ $(H_0 : \pi_b = \pi_w)$	$2.17_{(3,316)}$		0.0918
$F_{(df1,df2)}$ $(H_0 : \pi_b = 2 \cdot \pi_w)$	$1.97_{(3,316)}$		0.1183

[a]Columns 1 and 2 report coefficients associated with the least squares regression fit of test scores on the between- and within-classroom transforms of race (BLACK), gender (GIRL), and eligibility for free/reduced price school lunch (FREELUNCH); also included in the regression are school dummies and class type indicators (coefficients not reported). Column 1 reports estimated coefficients on the between-group transforms, column 2 reports coefficients on the within-group transforms, and column 3 reports the difference in these two sets of coefficients variable-by-variable. The * denotes the significance of these differences at the 5 percent level. Reported standard errors are heteroscedastic robust with clustering at the classroom level. The Omnibus Test Results panel reports F tests for the stated multicoefficient restriction along with degrees of freedom and asymptotic p-values.

Under positive social interactions the magnitude of the between-group coefficients (in absolute value) should be greater than the corresponding within-group coefficients. The omnibus test for no excess sensitivity is marginally accepted with p-values of 0.1304 and 0.0918 for math and reading test scores, respectively. The only individual-level covariate displaying significant excess sensitivity is gender (cf. Hoxby (2002)).

Overall, the excess sensitivity tests do not provide strong evidence of peer group effects. However, they also provide little evidence against the existence of even quite large effects. Table IV also reports tests for the restriction $\pi_{b0} = 2 \cdot \pi_{w0}$, which would hold if the true social multiplier were 2—a value similar to that implied by the estimates of γ^2 reported in Table I of the main paper. The test accepts with a p-value of 0.2027 for math achievement and marginally accepts with a p-value of 0.1183 for reading achievement. The excess sensitivity regressions are consistent with both very small and very large levels of peer group effects. The excess sensitivity test is uninformative.

C.1. *Power Comparisons*

To compare the relative merits of the excess sensitivity and variance approaches to testing for social interactions, it is useful to explicitly contrast their large sample power to reject the no social interactions null across repeated samples. In particular I consider samples drawn from a population calibrated to the Project STAR data set. While the comparison is necessarily design-specific, it is relevant to the application at hand. The chosen calibration both mimics the Project STAR data set and ensures that the excess sensitivity and excess variance tests are valid tests of social interactions.

Let R_{ci} be the vector of individual-level covariates used by the excess sensitivity estimator/test and let W_c be the binary instrument used by the excess variance estimator/test. I assume that repeated random samples of social groups of size N are drawn from

$$(18) \qquad \underline{Y}_c = R_c \eta + \overline{R}_c'(\gamma - 1)\eta + \underline{U}_c$$

with $R_c = (R_{c1}, \ldots, R_{cM_c})'$, $\overline{R}_c = R_c' \iota_{M_c}/M_c$, and

$$\underline{U}_c | R_c, M_c, W_c \sim \mathcal{N}(\underline{0}, \Omega(M_c)),$$
$$\Omega(M_c) = \sigma^2 \{ I_{M_c} + [\rho_\alpha + (\gamma^2 - 1)M_c^{-1}] \iota_{M_c} \iota_{M_c}' \},$$

where $\rho_\alpha = \sigma_\alpha^2/\sigma^2$ and $\sigma^2 = \sigma_\varepsilon^2 = \mathbb{V}(\varepsilon_{ci})$.

The marginal distribution of the individual-level covariates is also assumed to be normal with

$$R_{ci} | M_c, W_c \sim \mathcal{N}(\underline{0}, \Sigma_{RR}),$$

where the mean-zeroness assumption is without loss of generality. Integrating out the observed covariates, we get

$$(19) \qquad \underline{Y}_c | M_c, W_c \sim \mathcal{N}(\underline{0}, \Omega^*(M_c))$$

with

$$\Omega^*(M_c) = \sigma_*^2 \{ I_{M_c} + [\rho_\alpha^* + (\gamma^2 - 1) M_c^{-1}] \iota_{M_c} \iota'_{M_c} \},$$

where $\sigma_*^2 = \sigma^2 + \eta' \Sigma_{RR} \eta$ and $\rho_\alpha^* = \sigma_\alpha^2 / \sigma_*^2$.

This DGP is consistent with a linear-in-means model for outcomes with random assignment of individuals to groups. The auxiliary normality assumption is convenient, as some assumption on the fourth moments of α, R, and ε is required in order to calculate the asymptotic power function of the excess variance test. Normality is not exploited by either test. The presumption is that normality happens to be a feature of the population being sampled from, but that its does not inform estimation and testing procedures.

The derivation of the asymptotic power functions of the excess variance and sensitivity tests for the above DGP is straightforward (full details are provided below). Andrews' (1989) article is a standard reference for this type of analysis. Panel A of Table V gives the parameters used in the calibration. The sample size and distribution of group size are exactly as in the extract of the Project STAR data used in Section 3 of the main paper. The variance of unobserved individual heterogeneity parameter, σ^2, is the rounded sample mean of $M_c \cdot G_c^W$. The variance of observed heterogeneity is based on the estimates of η reported

TABLE V

APPROXIMATE POWER OF EXCESS SENSITIVITY AND VARIANCE TESTS[a]

Panel A			Panel B		
	Calibration Parameters			Excess Sensitivity	Excess Variance
σ^2	0.75	Power		0.8960	0.9994
σ_α^2	0.035	Inner inverse		1.5587	1.3922
$\eta' \Sigma_{RR} \eta$	0.05	Outer inverse		1.9645	1.6509
γ	$\sqrt{3.5} \approx 1.87$				
$K = \dim(R_{ci})$	3				
$F(M_c)$	As in STAR data				
N	317				

[a] Panel A reports the parameters that describe the calibrated population used for the power comparison. Panel B gives the large sample power of the two tests to reject the null of no social interactions across repeated samples drawn from the calibrated population. Panel B also reports the inner and outer inverse power functions for the two tests using the methods of Andrews (1989). The inner inverse equals the value of the social multiplier for which the given test fails to reject at least 50 percent of the time. In populations where $\gamma < \gamma^{\text{II}}$, the given test will be worse than one based on a coin flip. The outer inverse equals the value of γ for which the test rejects at least 95 percent of the time. In populations for which $\gamma > \gamma^{\text{OI}}$, the given test will be highly reliable.
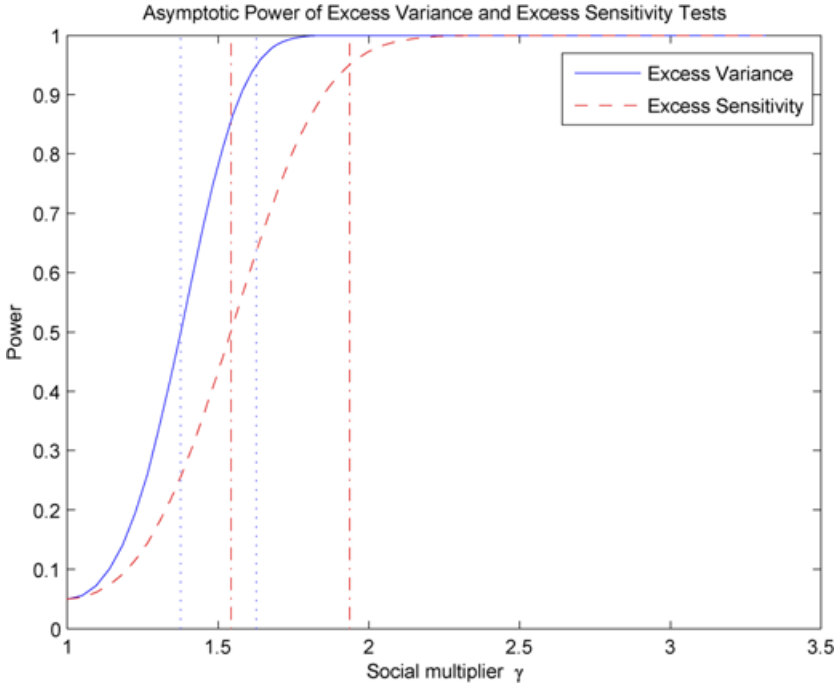
FIGURE 1.—Figure plots approximate (large sample) power functions for the excess variance and excess sensitivity tests of the no social interactions null. Vertical lines plot inner and outer inverse power envelopes as described by Andrews (1989). Auxiliary parameters are $\sigma^2 = 0.75$, $\sigma_\alpha^2 = 0.0035$, $\eta' \Sigma_{RR} \eta = 0.05$, and $K = \dim(R_{ci}) = 3$. The class size distribution and class type designations are as given by the Project STAR data.

in Table IV and the sample covariance matrix of R_{ci} (deviated from school means). The variance of group-level heterogeneity, σ_α^2, is somewhat speculatively set equal to 0.0035.[32] This implies a standard deviation in teaching quality of about 0.2 (i.e., that teachers are an important source of achievement variation).

The results suggest that the excess variance test is substantially more powerful than the excess sensitivity test across repeated samples drawn from the calibrated population. This power advantage is depicted visually in Figure 1. Panel B of Table V also reports the inner and outer inverse power envelopes for the two tests using the methods of Andrews (1989).

---

[32]In an earlier version of this paper (Graham (2005, Chapter 1, Table 17)), I applied a maximum likelihood estimator to a model equivalent to (19) using the Project STAR data. For the math data the (implied) point estimate of σ_α^2 was 0.0036. This value is somewhat higher than the (lower bound) estimates of the variance of teacher quality given by Rivkin, Hanushek, and Kain (2005, p. 434) based on middle school data from Texas.

These power comparisons are design-specific. It is, of course, possible to construct examples for which the excess sensitivity test is superior. A full characterization of the two power functions would raise a variety of issues that are beyond the scope of this appendix. The value of the calibration is that it helps to explain the strong evidence of social interactions provided by the excess variance estimator and the lack of evidence for such interactions provided by the excess sensitivity estimator. The contradictory test results appear to be simply an artifact of substantial differences in the design-specific power of the two tests.

C.2. *Details of Power Calculations*

This appendix provides details on the excess sensitivity and variance power calculations reported in Table V. The numerical calculations were completed using a short MATLAB program.

C.2.1. *Excess Sensitivity Test*

Rewriting (18) we have

$$\underline{Y}_c = X_c \pi + \underline{U}_c,$$

where $X_c = (\widetilde{R}_c, \iota_{M_c} \overline{R}'_c)$ and $\pi = (\eta', \eta\gamma')'$ with $\widetilde{R}_c = R_c - \iota_{M_c} \overline{R}'_c$. The large sample variance–covariance matrix associated with the least squares estimate of π is given by

$$\mathrm{AVar}(\widehat{\pi}) = \sigma^2 \mathbb{E}[X'_c X_c]^{-1} \mathbb{E}[X'_c \Omega(M_c) X_c] \mathbb{E}[X'_c X_c]^{-1}.$$

The $\mathbb{E}[X'_c X_c]$ term evaluates to

$$\mathbb{E}[X'_c X_c] = \mathbb{E}\begin{bmatrix} \sum_{i=1}^{M_c} \widetilde{R}_{ci} \widetilde{R}'_{ci} & 0 \\ 0 & M_c \overline{R}_c \overline{R}'_c \end{bmatrix} = \begin{pmatrix} \mu_M - 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \Sigma_{RR},$$

where $\mu_M = \mathbb{E}[M_c]$.

The $\mathbb{E}[X'_c \Omega(M_c) X_c]$ term simplifies as

$$\mathbb{E}[X'_c \Omega(M_c) X_c]$$

$$= \mathbb{E}\left[\sum_{i=1}^{M_c} X_{ci} X'_{ci} + (\rho_\alpha + (\gamma^2 - 1) M_c^{-1}) \sum_{i=1}^{M_c} \sum_{j=1}^{M_c} X_{ci} X'_{cj} \right]$$

$$= \begin{pmatrix} \mu_M - 1 & 0 \\ 0 & \gamma^2 + \mu_M \rho_\alpha \end{pmatrix} \otimes \Sigma_{RR}.$$

The large sample variance–covariance matrix for $\widehat{\pi}$ thus simplifies to

$$\text{AVar}(\widehat{\pi}) = \sigma^2 \left[\left(\begin{matrix} \frac{1}{\mu_M - 1} & 0 \\ 0 & \gamma^2 + \mu_M \rho_\alpha \end{matrix} \right) \otimes \Sigma_{RR}^{-1} \right],$$

while the asymptotic variance–covariance matrix of the difference $\widehat{\pi}_b - \widehat{\pi}_w$ is given by

$$(20) \qquad \text{AVar}(\widehat{\pi}_b - \widehat{\pi}_w) = \sigma^2 \left(\gamma^2 + \mu_M \rho_\alpha + \frac{1}{\mu_M - 1} \right) \cdot \Sigma_{RR}^{-1}.$$

To evaluate the power of the excess sensitivity test I use the standard Pitman drift approach. In particular I consider a sequence of alternative DGPs, where the social multiplier evolves with N such that

$$\gamma_N = 1 + \delta_0 / \sqrt{N}.$$

Observe that $\pi_b - \pi_w = \delta_0 \eta / \sqrt{N}$, which approaches zero as the sample size grows. The alternative DGP thus remains within a $1/\sqrt{N}$ neighborhood of the fixed no social interactions null. This prevents the asymptotic power function from taking a degenerate ⊤-shape since the scaled difference $\sqrt{N}(\gamma_N - 1)$ remains constant at δ_0.

Under this setup the scaled difference $\sqrt{N}(\widehat{\pi}_{b,N} - \widehat{\pi}_{w,N})$ converges in distribution to a normal random variable with mean $\delta_0 \eta$ and a variance given by (20) evaluated at the null of $\gamma = 1$. The excess sensitivity Wald statistic for the no social interactions null therefore converges in distribution to a noncentral $\chi^2_{K,\lambda}$ random variable with $\dim(R_{ci}) = K$ degrees of freedom and a noncentrality parameter of

$$\lambda = \eta' \delta_0 \, \text{AVar}(\widehat{\pi}_b - \widehat{\pi}_w)^{-1} \delta_0 \eta.$$

This implies that we can approximate the distribution of the Wald statistic for a given DGP (in the sequence) by a $\chi^2_{K,\lambda}$ random variable with a noncentrality parameter of

$$\lambda = N(\gamma - 1)^2 \frac{1}{\sigma^2} \left(1 + \mu_M \rho_\alpha + \frac{1}{\mu_M - 1} \right)^{-1} \cdot \eta' \Sigma_{RR} \eta.$$

This approximation is used for the power calculations reported in Table V and Figure 1.

C.2.2. *Excess Variance Test*

From (19) it is straightforward to show, using the normality assumption, that

$$G_c^w | M_c, W_c \sim \frac{\sigma_*^2}{M_c} \frac{1}{M_c - 1} \chi^2_{M_c - 1},$$

$$G_c^b | M_c, W_c \sim \left(\sigma_\alpha^2 + \frac{\gamma^2 \sigma_*^2}{M_c} \right) \chi_1^2.$$

Rewriting $\rho(Z_c, \theta) = G_c^b - \tau^2 - \gamma^2 G_c^w$ as

$$\rho(Z_c, \theta) = (G_c^b - \mathbb{E}[G_c^b | M_c]) - \gamma^2 (G_c^w - \mathbb{E}[G_c^w | M_c]),$$

using independence of G_c^b, and G_c^w and the properties of the χ^2 distribution, we then get

$$\mathbb{E}[G_c^w | M_c] = \frac{\sigma_*^2}{M_c}, \quad \mathbb{E}[G_c^b | M_c] = \sigma_\alpha^2 + \frac{\gamma^2 \sigma_*^2}{M_c},$$

and

$$(21) \quad \mathbb{E}[\rho(Z_c, \theta)^2 | W_c] \\ = 2\sigma_*^4 \{ (\rho_\alpha^* + \gamma^2 \mathbb{E}[M_c^{-1} | W_c])^2 + \gamma^4 \mathbb{E}[M_c^{-2}(M_c - 1)^{-1} | W_c] \}.$$

The unconditional moment function associated with the excess variance estimator for the binary instrument case is given by

$$\mathbb{E}[\psi(Z_c, \theta)] = \mathbb{E}\left[\begin{pmatrix} W_c \\ 1 - W_c \end{pmatrix} \rho(Z_c, \theta) \right] = 0.$$

The expected Jacobian matrix equals

$$(22) \quad \Gamma_0 = -\mathbb{E}\left[\begin{pmatrix} W_c \\ 1 - W_c \end{pmatrix} (1 \quad G_c^w) \right] \\ = -\begin{pmatrix} p & p\sigma_*^2 \mathbb{E}[M_c^{-1} | W_c = 1] \\ 1 - p & (1 - p)\sigma_*^2 \mathbb{E}[M_c^{-1} | W_c = 0] \end{pmatrix},$$

where $\mathbb{E}[W_c] = p$.

The variance of the moment vector, using iterated expectations, equals

$$(23) \quad \Lambda_0 = \mathbb{E}\left[\begin{pmatrix} p\mathbb{E}[\rho(Z_c, \theta)^2 | W_c = 1] & 0 \\ 0 & (1 - p)\mathbb{E}[\rho(Z_c, \theta)^2 | W_c = 0] \end{pmatrix} \right].$$

Standard GMM results yield a large sample variance–covariance matrix of $\text{AVar}(\widehat{\theta}) = (\Gamma_0' \Lambda_0^{-1} \Gamma_0)^{-1}$; the lower right-hand element of this matrix gives the large sample approximation to the sampling distribution of $\widehat{\gamma}^2$. Multiplying out, using (21), (22), and (23), and applying standard results on partitioned

inverses, we get

$$\frac{\text{AVar}(\widehat{\gamma}^2)}{N}$$

$$= \frac{1}{N} \frac{\frac{2}{p}(\rho_\alpha^* + \gamma^2 \mathbb{E}[M_c^{-1}|W_c=1])^2 + \frac{2}{1-p}(\rho_\alpha^* + \gamma^2 \mathbb{E}[M_c^{-1}|W_c=0])^2}{(\mathbb{E}[M_c^{-1}|W_c=1] - \mathbb{E}[M_c^{-1}|W_c=0])^2}$$

$$+ \frac{\gamma^4}{\kappa_0},$$

where

$$\kappa_0 = \frac{N}{2} \frac{(\mathbb{E}[M_c^{-1}|W_c=1] - \mathbb{E}[M_c^{-1}|W_c=0])^2}{\frac{1}{p}\mathbb{E}[M_c^{-2}(M_c-1)^{-1}|W_c=1] + \frac{1}{1-p}\mathbb{E}[M_c^{-2}(M_c-1)^{-1}|W_c=0]}$$

is the concentration parameter associated with the first stage regression of G_c^w on a constant and W_c (cf. Staiger and Stock (1997)).

To evaluate the power properties of the excess variance test, I consider the sequence of alternative DGPs with $\gamma_N^2 = 1 + \delta_0/\sqrt{N}$. Using an argument entirely analogous to the excess variance case, we can then approximate the sampling distribution of the excess variance Wald statistic for a specific DGP in the sequence with that of a $\chi_{1,\lambda}^2$ random variable with 1 degree of freedom and a noncentrality parameter of

$$\lambda = (\gamma^2 - 1)^2$$

$$\times \left(\frac{1}{N} \frac{\frac{2}{p}(\rho_\alpha^* + \mathbb{E}[M_c^{-1}|W_c=1])^2 + \frac{2}{1-p}(\rho_\alpha^* + \mathbb{E}[M_c^{-1}|W_c=0])^2}{(\mathbb{E}[M_c^{-1}|W_c=1] - \mathbb{E}[M_c^{-1}|W_c=0])^2} \right.$$

$$\left. + \frac{1}{\kappa_0} \right)^{-1}.$$

D. DATA APPENDIX

The core data used for this paper are from the *Project STAR K-3 Public Access Dataset* available online in a variety of machine readable formats at http://www.heros-inc.org/data.htm. A STATA formatted version of these data as well as STATA Do files replicating the extraction used in the paper as well the reported estimation results is available online at http://www.econ.berkeley.edu/~bgraham/ as well as in the Supplemental Materials section of the Econometric Society website. The articles by Krueger (1999) and Finn, Gerber, Achilles, and Boyd-Zaharias (2001) provide a nice overview of the public release data.

The public release Project STAR data set do not include a classroom identi-fier. However, using a simple algorithm based on grouping students with common values for school, class type (small, regular, or regular-with-aide), and teacher characteristics, I was able to uniquely assign 6,172 students to 317 classrooms; this is the sample used in the paper. Boozer and Cacciola (2004) used a similar algorithm. Of the eight kindergarten classrooms excluded from the analysis, two are regular classrooms and four are small classrooms which could not be individually separated; a further two classrooms were missing some teacher data and were also dropped. Twenty-three kindergarten student records were missing information on free and reduced price school lunch el-igibility; in these cases the missing values were replaced with either eligibility status for the same student in the closest of first, second, or third grade (17 cases) or the median value among kindergarten students in their school (6 cases). In three cases, missing student race values were replaced with school median values.

Valid test scores are not available for all kindergarten students. For the math test, 5,724 students have valid scores and for the reading test, 5,646 scores are valid (out of the 6,172 students in the core sample described above). Omis-sions of test scores appear to be idiosyncratic, in the sense that they are not predictable by any observable student, teacher, or peer covariates. The analy-sis below assumes the pattern of missing test score data is indeed completely random (see the log files associated with the Do files referenced above). The definitions of G_c^w and G_c^b are modified as described in Appendix F.

E. DETAILED DERIVATIONS OF EQUATIONS (6) AND (8) IN THE MAIN PAPER

This appendix details the calculations used to derive equations (6) and (8) in the main paper. Recall the notation

$$\mathbb{V}(\alpha_c|m,w) = \sigma_\alpha^2(m,w), \quad \mathbb{V}(\varepsilon_{ci}|m,w) = \sigma^2(m,w),$$
$$\mathbb{C}(\alpha_c, \varepsilon_{ci}|m,w) = \sigma_{\alpha\varepsilon}(m,w), \quad \mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|m,w) = \sigma_{\varepsilon\varepsilon}(m,w),$$
$$\lambda^2(m,w) = \sigma^2(m,w) - \sigma_{\varepsilon\varepsilon}(m,w),$$
$$\tau_0^2(m,w) = \sigma_\alpha^2(m,w) + 2\gamma_0\sigma_{\alpha\varepsilon}(m,w) + \gamma_0^2\sigma_{\varepsilon\varepsilon}(m,w).$$

Since the $M_c \times M_c$ conditional covariance matrix of outcomes $\Omega(m,w)$ has an equicorrelated structure, we need only evaluate $\mathbb{V}(Y_{ci}|m,w)$ and $\mathbb{C}(Y_{ci}, Y_{cj}|m,w)$ to derive equation (6).

We have

$$\mathbb{V}(Y_{ci}|m,w) = \mathbb{V}(\alpha_c|m,w) + (\gamma_0-1)^2\mathbb{V}(\bar{\varepsilon}_c|m,w) + \mathbb{V}(\varepsilon_{ci}|m,w)$$
$$+ 2(\gamma_0-1)\mathbb{C}(\alpha_c, \bar{\varepsilon}_c|m,w) + 2\mathbb{C}(\alpha_c, \varepsilon_{ci}|m,w)$$
$$+ 2(\gamma_0-1)\mathbb{C}(\bar{\varepsilon}_c, \varepsilon_{ci}|m,w).$$

Now observe that

$$
\mathbb{V}(\bar{\varepsilon}_c|m, w) = \mathbb{V}\left(\frac{1}{M_c}\sum_{i=1}^{M_c}\varepsilon_{ci}\Big|m, w\right)
$$

$$
= \frac{1}{m^2}\left[\sum_{i=1}^{m}\mathbb{V}(\varepsilon_{ci}|m, w) + \sum_{i=1}^{m}\sum_{j\neq i}\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|m, w)\right]
$$

$$
= \frac{1}{m^2}\left[\sum_{i=1}^{m}\sigma^2(m, w) + \sum_{i=1}^{m}\sum_{j\neq i}\sigma_{\varepsilon\varepsilon}(m, w)\right]
$$

$$
= \frac{\sigma^2(m, w)}{m} + \frac{m-1}{m}\sigma_{\varepsilon\varepsilon}(m, w)
$$

$$
= \sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}.
$$

By linearity of the expectations operator, we have

$$
\mathbb{C}(\alpha_c, \bar{\varepsilon}_c|m, w) = \frac{1}{m}\sum_{i=1}^{m}\mathbb{C}(\alpha_c, \varepsilon_{ci}|m, w) = \sigma_{\alpha\varepsilon}(m, w)
$$

and also that

$$
\mathbb{C}(\bar{\varepsilon}_c, \varepsilon_{ci}|m, w) = \frac{1}{m}\sum_{j=1}^{m}\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|m, w)
$$

$$
= \frac{1}{m}[\sigma^2(m, w) + (m-1)\sigma_{\varepsilon\varepsilon}(m, w)]
$$

$$
= \sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}.
$$

Using these expressions we therefore have

$$
\mathbb{V}(Y_{ci}|m, w)
$$

$$
= \sigma_\alpha^2(m, w) + (\gamma_0 - 1)^2\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right]
$$

$$
+ \sigma^2(m, w) + 2(\gamma_0 - 1)\sigma_{\alpha\varepsilon}(m, w) + 2\sigma_{\alpha\varepsilon}(m, w)
$$

$$
+ 2(\gamma_0 - 1)\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right].
$$

Adding and subtracting $\sigma_{\varepsilon\varepsilon}(m, w) + (\lambda^2(m, w))/m$, factoring, and rearranging then give

$$
\begin{aligned}
\mathbb{V}(Y_{ci}|m, w) \\
&= \sigma_\alpha^2(m, w) + 2\gamma_0\sigma_{\alpha\varepsilon}(m, w) + \gamma_0^2\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] \\
&\quad + \sigma^2(m, w) - \left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] \\
&= \sigma_\alpha^2(m, w) + 2\gamma_0\sigma_{\alpha\varepsilon}(m, w) \\
&\quad + (\gamma_0^2 - 1)\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] + \sigma^2(m, w) \\
&= \sigma_\alpha^2(m, w) + 2\gamma_0\sigma_{\alpha\varepsilon}(m, w) + \gamma_0^2\sigma_{\varepsilon\varepsilon}(m, w) \\
&\quad + (\gamma_0^2 - 1)\frac{\lambda^2(m, w)}{m} + \sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w) \\
&= \lambda^2(m, w) + \tau_0^2(m, w) + (\gamma_0^2 - 1)\frac{\lambda^2(m, w)}{m},
\end{aligned}
$$

as given in equation (6) of the main text.

We can evaluate the covariance terms analogously. We have

$$
\begin{aligned}
\mathbb{C}(Y_{ci}, Y_{cj}|m, w) \\
&= \mathbb{V}(\alpha_c|m, w) + (\gamma_0 - 1)^2\mathbb{V}(\overline{\varepsilon}_c|m, w) + \mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|m, w) \\
&\quad + 2(\gamma_0 - 1)\mathbb{C}(\alpha_c, \overline{\varepsilon}_c|m, w) + \mathbb{C}(\alpha_c, \varepsilon_{ci}|m, w) + \mathbb{C}(\alpha_c, \varepsilon_{cj}|m, w) \\
&\quad + (\gamma_0 - 1)\mathbb{C}(\overline{\varepsilon}_c, \varepsilon_{ci}|m, w) + (\gamma_0 - 1)\mathbb{C}(\overline{\varepsilon}_c, \varepsilon_{cj}|m, w) \\
&= \sigma_\alpha^2(m, w) + (\gamma_0 - 1)^2\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] + \sigma_{\varepsilon\varepsilon}(m, w) \\
&\quad + 2(\gamma_0 - 1)\sigma_{\alpha\varepsilon}(m, w) + 2\sigma_{\alpha\varepsilon}(m, w) \\
&\quad + 2(\gamma_0 - 1)\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] \\
&= \sigma_\alpha^2(m, w) + 2\gamma_0\sigma_{\alpha\varepsilon}(m, w) + \sigma_{\varepsilon\varepsilon}(m, w) \\
&\quad + (\gamma_0^2 - 1)\left[\sigma_{\varepsilon\varepsilon}(m, w) + \frac{\lambda^2(m, w)}{m}\right] \\
&= \tau_0^2(m, w) + (\gamma_0^2 - 1)\frac{\lambda^2(m, w)}{m},
\end{aligned}
$$

as given in equation (6) of the main text.

Now consider the expectations of the within- and between-group transforms

$$G_c^w = \frac{1}{M_c} \frac{1}{M_c - 1} \sum_{i=1}^{M_c} (Y_{ci} - \overline{Y}_c)^2, \quad G_c^b = (\overline{Y}_c - \mu_0(W_c))^2.$$

The expectations given in equation (8) of the main text follow directly from the derivation of $\Omega(m, w)$ given above. For completeness, the details are given here.

Observe that

$$\mathbb{E}[G_c^w | W_c] = \mathbb{E}\big[\mathbb{E}[G_c^w | M_c, W_c] | W_c\big].$$

Evaluating $\mathbb{E}[G_c^w | m, w]$ gives

$$\mathbb{E}[G_c^w | m, w] = \frac{1}{m} \frac{1}{m - 1} \sum_{i=1}^{m} \mathbb{E}[(\overline{\varepsilon}_c - \varepsilon_{ci})^2 | m, w]$$

$$= \frac{1}{m} \frac{1}{m - 1} \sum_{i=1}^{m} \mathbb{E}[\overline{\varepsilon}_c^2 - 2\overline{\varepsilon}_c \varepsilon_{ci} + \varepsilon_{ci}^2 | m, w]$$

$$= \frac{1}{m} \frac{1}{m - 1} \sum_{i=1}^{m} \mathbb{E}[\overline{\varepsilon}_c^2 - 2\overline{\varepsilon}_c \varepsilon_{ci} + \varepsilon_{ci}^2 | m, w]$$

$$= \frac{1}{m} \frac{m}{m - 1} \left[\begin{array}{c} \frac{\sigma^2(m,w)}{m} + \frac{m-1}{m} \sigma_{\varepsilon\varepsilon}(m, w) \\ - \frac{2\sigma^2(m,w) + (m-1)\sigma_{\varepsilon\varepsilon}(m,w)}{m} + \sigma^2(m, w) \end{array} \right]$$

$$= \frac{1}{m - 1} \left[\sigma^2(m, w) - \frac{\sigma^2(m, w)}{m} - \frac{m - 1}{m} \sigma_{\varepsilon\varepsilon}(m, w) \right]$$

$$= \frac{1}{m - 1} \left[\frac{m - 1}{m} \sigma^2(m, w) - \frac{m - 1}{m} \sigma_{\varepsilon\varepsilon}(m, w) \right]$$

$$= \frac{1}{m - 1} \left[\frac{m - 1}{m} \sigma^2(m, w) - \frac{m - 1}{m} \sigma_{\varepsilon\varepsilon}(m, w) \right]$$

$$= \frac{\lambda^2(m, w)}{m},$$

where we use conditional mean-zeroness of ε_{ci} and the expression for $\mathbb{V}(\overline{\varepsilon}_c | m, w)$ derived above. Taking expectations then gives the first part of equation (8) in the main text.

Evaluating the between-group square we have

$$\mathbb{E}[G_c^b | w] = \mathbb{E}\big[(\overline{Y}_c - \mu_0(W_c))^2 | w\big]$$

$$= \mathbb{V}(\overline{Y}_c|w)$$
$$= \mathbb{V}(\alpha_c|w) + \gamma_0^2 \mathbb{V}(\overline{\varepsilon}|w) + 2\gamma_0 \mathbb{V}(\alpha_c, \overline{\varepsilon}|w)$$
$$= \sigma_\alpha^2(w) + \gamma_0^2 \mathbb{V}(\overline{\varepsilon}|w) + 2\gamma_0 \mathbb{C}(\alpha_c \overline{\varepsilon}|w).$$

Now use the analysis of variance formula to decompose $\mathbb{V}(\overline{\varepsilon}|w)$, $\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|w)$, and $\mathbb{C}(\alpha_c, \overline{\varepsilon}|w)$ as

$$\mathbb{V}(\overline{\varepsilon}|w) = \mathbb{E}[\mathbb{V}(\overline{\varepsilon}|M_c, W_c)|w] + \mathbb{V}(\mathbb{E}[\overline{\varepsilon}|M_c, W_c]|w)$$
$$= \mathbb{E}\left[\sigma_{\varepsilon\varepsilon}(M_c, W_c) + \frac{\lambda^2(M_c, W_c)}{M_c}\Big|w\right] + 0,$$

where conditional mean-zeroness of ε_{ci} is used. We also have

$$\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|w) = \mathbb{E}[\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj}|M_c, W_c)|w] + 0 = \sigma_{\varepsilon\varepsilon}(w),$$
$$\mathbb{C}(\alpha_c, \overline{\varepsilon}|w) = \mathbb{E}[\mathbb{C}(\alpha_c, \overline{\varepsilon}|M_c, W_c)|w] + 0 = \sigma_{\alpha\varepsilon}(w).$$

Collecting terms then gives the desired result:

$$\mathbb{E}[G_c^b|w] = \sigma_\alpha^2(w) + 2\gamma_0 \sigma_{\alpha\varepsilon}^2(w) + \gamma_0^2 \sigma_{\varepsilon\varepsilon}(w) + \gamma_0^2 \mathbb{E}\left[\frac{\lambda^2(M_c, W_c)}{M_c}\Big|w\right]$$
$$= \tau_0^2(w) + \gamma_0^2 \mathbb{E}\left[\frac{\lambda^2(M_c, W_c)}{M_c}\Big|w\right].$$

F. FORMS OF G_c^w AND G_c^b WHEN ONLY A RANDOM SUBSAMPLE OF INDIVIDUALS IN EACH GROUP IS OBSERVED

With minor modification the identification results of Section 1 remain valid if the econometrician only observes outcomes for a random subsample of group members. Let $M_c^S \leq M_c$ denote the number of sampled individuals in the cth group. Assume that the M_c^S sampled individuals are a random subsample of all M_c group members and let $\overline{Y}_c^S = \sum_{i=1}^{M_c^S} Y_{ci}/M_c^S$ equal their mean outcome. Redefine G_c^w and G_c^b to equal

$$G_c^b = (\overline{Y}_c^S - \mu_Y(W_c))^2 - \left(\frac{1}{M_c^S} - \frac{1}{M_c}\right)\frac{1}{M_c^S - 1}\sum_{i=1}^{M_c^S}(Y_{ci} - \overline{Y}_c^S)^2,$$

$$G_c^w = \frac{1}{M_c}\frac{1}{M_c^S - 1}\sum_{i=1}^{M_c^S}(Y_{ci} - \overline{Y}_c^S)^2.$$

Subject to the above redefinition of G_c^w and G_c^b, the estimators discussed in the main text remain appropriate when outcomes for only a random subsample of all group members are observed.

To verify this claim observe that $\overline{Y}_c^S = \alpha_c + (\gamma_0 - 1)\overline{\varepsilon}_c + \overline{\varepsilon}_c^S$. This implies that the between-group variance in *observed* outcomes is given by

$$\mathbb{V}(\overline{Y}_c^S | m, m^S, w)$$
$$= \mathbb{V}(\alpha_c | m, m^S, w) + (\gamma_0 - 1)^2 \mathbb{V}(\overline{\varepsilon}_c | m, m^S, w) + \mathbb{V}(\overline{\varepsilon}_c^S | m, m^S, w)$$
$$+ 2(\gamma_0 - 1)\mathbb{C}(\alpha_c, \overline{\varepsilon}_c | m, m^S, w) + 2\mathbb{C}(\alpha_c, \overline{\varepsilon}_c^S | m, m^S, w)$$
$$+ 2(\gamma_0 - 1)\mathbb{C}(\overline{\varepsilon}_c, \overline{\varepsilon}_c^S | m, m^S, w).$$

To simplify this expression observe that since sampled group members are a random subsample of all group members, we have the restrictions

$$\mathbb{V}(\varepsilon_{ci} | m, m^S, w) = \mathbb{V}(\varepsilon_{ci} | m, w),$$
$$\mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj} | m, m^S, w) = \mathbb{C}(\varepsilon_{ci}, \varepsilon_{cj} | m, w),$$
$$\mathbb{C}(\alpha_c, \varepsilon_{ci} | m, m^S, w) = \mathbb{C}(\alpha_c, \varepsilon_{ci} | m, w).$$

Now verify the equality

$$\mathbb{C}(\overline{\varepsilon}_c, \overline{\varepsilon}_c^S | m, m^S, w)$$
$$= \mathbb{E}\left[\left(\frac{1}{M_c^S}\sum_{i=1}^{M_c^S}\varepsilon_{ci}\right)\left(\frac{1}{M_c}\sum_{i=1}^{M_c}\varepsilon_{ci}\right)\middle| m, m^S, w\right]$$
$$= \frac{1}{m^S}\frac{1}{m}\sum_{i=1}^{m^S}\mathbb{E}[\varepsilon_{ci}^2 | m, w] + \frac{1}{m^S}\frac{1}{m}\sum_{i=1}^{M_c^S}\sum_{j\neq i, j=1}^{M_c}\mathbb{E}[\varepsilon_{ci}\varepsilon_{cj} | m, w]$$
$$= \frac{\sigma^2(m, w)}{m} + \frac{m-1}{m}\sigma_{\varepsilon\varepsilon}(m, w)$$
$$= \mathbb{V}(\overline{\varepsilon}_c | m, w)$$

and also observe that

$$\mathbb{E}\left[\frac{1}{M_c^S - 1}\sum_{i=1}^{M_c^S}(\varepsilon_{ci} - \overline{\varepsilon}_c^S)^2 \middle| m, m^S, w\right] = \sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w).$$

Using these results and those given in Appendix E above implies that

$$\mathbb{V}(\overline{Y}_c^S | m, m^S, w)$$
$$= \sigma_\alpha^2(m, m^S, w) + (\gamma_0 - 1)^2\left[\frac{\sigma^2(m, w)}{m} + \frac{m-1}{m}\sigma_{\varepsilon\varepsilon}(m, w)\right]$$

$$+ \left[\frac{\sigma^2(m, w)}{m^S} + \frac{m^S - 1}{m^S} \sigma_{\varepsilon\varepsilon}(m, w) \right] + 2(\gamma_0 - 1)\sigma_{\alpha\varepsilon}(m, w)$$

$$+ 2\sigma_{\alpha\varepsilon}(m, w) + 2(\gamma_0 - 1) \left[\frac{\sigma^2(m, w)}{m} + \frac{m - 1}{m} \sigma_{\varepsilon\varepsilon}(m, w) \right]$$

$$= \sigma_\alpha^2(m, m^S, w) + 2\gamma_0 \sigma_{\alpha\varepsilon}(m, w)$$

$$+ [(\gamma_0 - 1)^2 + 2(\gamma_0 - 1) + 1]\sigma_{\varepsilon\varepsilon}(m, w)$$

$$+ (\gamma_0 - 1)^2 \left[\frac{\sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w)}{m} \right] + 2(\gamma_0 - 1)$$

$$\times \left[\frac{\sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w)}{m} \right] + \left[\frac{\sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w)}{m^S} \right]$$

$$= \sigma_\alpha^2(m, m^S, w) + 2\gamma_0 \sigma_{\alpha\varepsilon}(m, w) + \gamma_0^2 \sigma_{\varepsilon\varepsilon}(m, w)$$

$$+ \gamma_0^2 \left[\frac{\sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w)}{m} \right]$$

$$+ \left(\frac{1}{m^S} - \frac{1}{m} \right) [\sigma^2(m, w) - \sigma_{\varepsilon\varepsilon}(m, w)].$$

Observing that

$$\mathbb{V}(\overline{Y}_c^S | w) = \mathbb{E}[\text{Var}(\overline{Y}_c^S | m, m^S, w) | w] + \text{Var}(\mathbb{E}[\overline{Y}_c^S | m, m^S, w] | w)$$

and using the conditional mean-zeroness of ε_{ci} implies that

$$\mathbb{E}[G_c^b | w] = \tau_0^2(w) + \gamma_0^2 \mathbb{E}\left[\frac{\lambda^2(M_c, W_c)}{M_c} \Big| w \right]$$

and also that

$$\mathbb{E}[G_c^w | w] = \mathbb{E}\left[\frac{\lambda^2(M_c, W_c)}{M_c} \Big| w \right].$$

Using these expression is straightforward to verify that Proposition 1.1 remains valid.

G. ESTIMATION OF $\widehat{\gamma}^2$

Given a random sample, it is straightforward to estimate γ_0^2 using restriction (11). Feasible estimation requires replacing G_c^b in $\rho(Z_c, \theta)$ with the estimate $\widehat{G}_c^b = (\overline{Y}_c - \widehat{\mu}_Y(W_c))^2$, where $\widehat{\mu}_Y(W_c)$ is a consistent estimate of $\mu_Y(W_c)$. This feasible estimator has the same asymptotic variance as the infeasible estimator which uses G_c^b; standard errors do not need to be adjusted for sampling

error in $\widehat{\mu}_Y(W_c)$. When (12) holds with $\tau_0^2(W_{1c})$ an unknown smooth function, the sieve minimum distance estimator of Ai and Chen (2003) would be appropriate.

The argument for the claim that sampling error in $\widehat{\mu}_Y(W_c)$ need not be accounted for follows from the results of Newey (1994). To make the dependence of $\rho(Z, \theta)$ on the nuisance parameter $\mu_Y(W_c)$ explicit, we write

$$\rho(Z, \theta, \mu_Y(W)) = (\overline{Y} - \mu_Y(W))^2 - \tau^2 - \gamma^2 G_c^w.$$

For estimation we use the unconditional moment

$$\psi(Z, \theta, \mu_Y(W)) = A(W)\rho(Z, \theta, \mu_Y(W)).$$

Note that this moment only depends on the function μ_Y through its value $\mu_Y(W)$, where W is a subvector of Z. Differentiating with respect to μ_Y and evaluating at $\mu_Y = \mu_Y(W)$, we have

$$D(Z) = \left.\frac{\partial \psi(Z, \theta, \mu_Y)}{\partial \mu_Y}\right|_{\mu_Y = \mu_Y(W)} = -A(W)2(\overline{Y} - \mu_Y(W)),$$

and hence that $\mathbb{E}[D(Z)|W] = 0$, which, as shown by Newey (1994, Proposition 3, pp. 1359–1360), is a sufficient condition for the asymptotic sampling behavior of $\psi(Z, \theta, \widehat{\mu}_Y(W))$ to be the same as that of $\psi(Z, \theta, \mu_Y(W))$. Therefore sampling error in $\widehat{\mu}_Y(W)$ does not affect the asymptotic variance of $\widehat{\theta}$.

Assumptions 1.1–1.3, combined with additional auxiliary assumptions, can also be used to form consistent estimators for the parameters characterizing specific members of the family given by (13). Unfortunately, but not surprisingly, identification arguments are model-specific. To illustrate some of the issues involved, assume we are interested in estimating γ_0 when $h(\varepsilon_{ci}, \underline{\varepsilon}_c) = \min\{\underline{\varepsilon}_c\}$. We begin by augmenting Assumptions 1.1–1.3 with the parametric assumption that

$$(24) \quad \begin{pmatrix} \alpha_c \\ \underline{\varepsilon}_c \end{pmatrix} \Big| W_c$$
$$\sim N\left(\begin{pmatrix} \mu_\alpha(W_c) \\ 0 \end{pmatrix}, \begin{pmatrix} \exp(W_{1c}'\pi_\alpha) & 0 \\ 0 & \exp(W_{1c}'\pi_{1\varepsilon} + W_{2c}'\pi_{2\varepsilon})I_{Mc} \end{pmatrix} \right)$$

with $\pi_{2\varepsilon} \neq 0$. Note that (24) is sufficient to satisfy Assumptions 1.1–1.3. In this case, $\mu_a(W_c)$ and $\pi_\varepsilon = (\pi_{1\varepsilon}', \pi_{2\varepsilon}')'$ are identified by

$$\mathbb{E}[Y_{ci}|W_c] = \mu_\alpha(W_c) \quad \text{and} \quad \mathbb{E}[M_c \cdot G_c^w|W_c] = \exp(W_{1c}'\pi_{1\varepsilon} + W_{2c}'\pi_{2\varepsilon}).$$

Using the normality assumption, the conditional distribution of $\underline{\varepsilon}_c$ given W_c is therefore identified. We also have

$$\mathbb{E}[G_c^b - G_c^w | W_c; \pi_\alpha, \pi_\varepsilon, \gamma_0]$$
$$= \exp(W_{1c}' \pi_\alpha) + (\gamma_0 - 1)^2 v(W_c; \pi_\varepsilon) + 2(\gamma_0 - 1)c(W_c; \pi_\varepsilon),$$

where $v(W_c; \pi_\varepsilon) = \mathbb{V}(\min(\underline{\varepsilon}_c)|W_c; \pi_\varepsilon)$ and $c(W_c; \pi_\varepsilon) = \mathbb{C}(\min(\underline{\varepsilon}_c), \overline{\varepsilon}_c | W_c; \pi_\varepsilon)$.

With the conditional distribution of $\underline{\varepsilon}_c$ identified, we can compute $v(W_c; \pi_\varepsilon)$ and $c(W_c; \pi_\varepsilon)$ (perhaps by simulation). With $v(W_c; \pi_\varepsilon)$ and $c(W_c; \pi_\varepsilon)$ known, γ_0 is then identified by the nonlinear (in π_α and γ_0) regression function $\mathbb{E}[G_c^b - G_c^w | W_c]$. This sequential procedure can be given a method-of-moments representation, allowing for inference. The important point of the example, however, is that identification via conditional variance restrictions is not specific to the linear-in-means model of social interactions. Researchers with different focal models can apply the same basic identifying principles to estimation.

Dept. of Economics, University of California–Berkeley, 544 Evans Hull 3880, Berkeley, CA 94720-3880, U.S.A. and National Bureau of Economic Research; bgraham@econ.berkeley.edu; http://www.econ.berkeley.edu/~bgraham/.

REFERENCES

AARONSON, D., L. BARROW, AND W. SANDERS (2003): "Teachers and Student Achievement in Chicago Public High Schools," WP 2002-28, Federal Reserve Bank of Chicago. [4]

AI, C., AND X. CHEN (2003): "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843. [21]

ANDREWS, D. W. K. (1989): "Power in Econometric Applications," *Econometrica*, 57, 1059–1090. [8,9]

ANGRIST, J. D., AND K. LANG (2004): "Does School Integration Generate Peer Effects? Evidence From Boston's Metco Program," *American Economic Review*, 94, 1613–1634. [5]

BETTS, J. R., AND J. L. SHKOLNIK (1999): "The Behavioral Effects of Variation in Class Size: The Case of Math Teachers," *Educational Evaluation and Policy Analysis*, 21, 193–213. [3]

BOOZER, M. A., AND S. E. CACCIOLA (2004): "Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data," Mimeo, Yale University. [14]

EVERTSON, C. M., AND C. H. RANDOLPH (1989): "Teaching Practices and Class Size: A New Look at an Old Issue," *Peabody Journal of Education*, 67, 85–105. [3]

FINN, J. D., S. B. GERBER, C. M. ACHILLES, AND J. BOYD-ZAHARIAS (2001): "The Enduring Effects of Small Classes," *Teachers College Record*, 103, 145–183. [13]

GRAHAM, B. S. (2005): "Essavs on the Econometrics of Social Interactions," Ph.D. dissertation, Harvard University. [9]

GRAHAM, B. S., AND J. HAHN (2005): "Identification and Estimation of the Linear-in-Means Model of Social Interactions," *Economic Letters*, 88, 1–6. [5,6]

HAUSMAN, J. A., AND W. E. TAYLOR (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49, 1377–1398. [6]

HOXBY, C. M. (2002): "The Power of Peers: How Does the Makeup of a Classroom Influence Achievement?" *Education Next*, 2, 57–63. [7]

KRUEGER, A. B. (1999): "Experimental Estimates of Education Production Functions," *Quarterly Journal of Economics*, 114, 497–532. [13]

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. [21]

RICE, J. K. (1999): "The Impact of Class Size on Instructional Strategies and the Use of Time in High School Mathematics and Science Courses," *Educational Evaluation and Policy Analysis*, 21, 215–229. [3]

RIVKIN, S. G., E. A. HANUSHEK, AND J. F. KAIN (2005): "Teachers, Schools and Academic achievement," *Econometrica*, 73, 417–458. [4,9]

ROCKOFF, J. (2004): "The Impact of Individual Teachers on Student Achievement: Evidence From Panel Data," *American Economic Review*, 94, 247–252. [4]

SACERDOTE, B. (2001): "Peer Effects With Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116, 681–704. [5]

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586. [13]

WORD, E., J. JOHNSTON, H. P. BAIN, B. DEWAYNE FULTON, C. M. ACHILLES, M. N. LINTZ, J. FOLGER, AND C. BREDA (1990): "The State of Tennessee's Student/Teacher Ratio (STAR) Project," Technical Report 1985–1990, Mimeo. [4]