assumptions of value-added models for estimating school effects

sean f reardon

stephen w raudenbush

april, 2008

goals of this paper

- frame value-added models (VAMs) within counterfactual model of causality
- articulate six assumptions necessary to interpret estimated VAM parameters as causal effects
- discuss plausibility of these assumptions
- assess sensitivity of VAM estimates to violations of three of these assumptions

counterfactual model of school effects

- define the potential outcome, Y_i^j, as the test score for student i that would be observed (at some time t) if student i were assigned to school j.
- define G_P^j as the distribution of potential outcomes if all students in population P are assigned to school j.
- comparing effectiveness of schools j and k requires a comparison of G_P^j and G_P^k (typically we compare their means).

VAM assumptions (1) and (2)

- (1) manipulability: each student can, in principle, be assigned to any school j
  - must be independently manipulable
  - necessary in order to define potential outcome Y_i^j
- (2) no interference across units:
  - also known as stable unit treatment value assumption (SUTVA)
  - potential outcome Y_i^j must be unique; must not depend on assignment of other students
- both are required so that we can write down the estimand of interest

a stylized VAM

- consider a stylized VAM:

$$Y_i^j = f(\boldsymbol{x_i}) + \Delta_j + \epsilon_i^j,$$

$$\text{where } \epsilon_i^j \perp x_i, \Delta_j \,\&\, E\left[\epsilon_i^j | x_i, \Delta_j\right] = 0$$

- the mean potential outcome is then

$$\mu(G_P^j) = E\left[y_i^j | i \in P, j\right] = E\left[f(\boldsymbol{x_i}) | i \in P\right] + \Delta_j$$

- so we can compare schools j and k by comparing

$$\mu(G_P^j) - \mu(G_P^k) = \Delta_j - \Delta_k$$

- this relies on two additional assumptions

VAM assumptions (3) and (4)

- (3) homogeneity: average effect of school j is constant over **x**.
  - requires that if school j is better than school k for students with $\mathbf{x}=\mathbf{x}_1$, then school j is also better than school k for students with $\mathbf{x}=\mathbf{x}_2$. (a good school is good for every type of student)
- (4) interval metric: Y is measured in an interval-scaled metric
  - ranking schools by their mean potential outcomes implies Y is interval-scaled

fitting VAMs from observed data

- we cannot observe the full distribution of potential outcomes for all students in all schools
  - (this is "the fundamental problem of causal inference" Holland, 1986)
  - we must estimate the parameters of a VAM (especially Δ_j) from observed data
  - we use the observed/realized potential outcomes in school j (and their covariance with **x**) to estimate the parameter(s) of G_P^j (such as the mean of G_P^j).
  - this requires two additional assumptions

VAM assumptions (5) and (6)

- (5) ignorability: school assignment is independent of potential outcomes, given observed covariates **x**.
  - necessary in order to obtain unbiased estimates
- (6) common support/correct functional form: the functional form of the model is valid in regions where data are sparse
  - if there are no/few students with **x**=x in school j (no common support), then we rely on the functional form of the VAM and extrapolation to infer the mean potential outcomes of students with **x**=x in school j.

how plausible are these assumptions?

- manipulability
- no interference between units
- homogeneity
- interval metric
- ignorability
- common support/functional form

how sensitive are VAM estimates to violations of these assumptions?

- for the sake of argument, we assume manipulability, SUTVA, and ignorability
- we focus on implications of plausible violations of homogeneity, interval metric, and common support assumptions
  - assess sensitivity of estimates to violations of the assumptions via simulation exercises

simulation study

- 'true' model for potential outcomes (data generating model):

$$Y_i^j = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})(Y_{i0} - \bar{\bar{Y}}_0) + e_{ij}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

  - this is a very simple model (too simple)
  - heterogeneity parameter is τ_{11}
    - we use $\boldsymbol{\tau_{11} = 0, 0.2, 0.4}$

observed data parameters

- given the true potential outcomes, two key parameters determine features of the observed data:
  - intracluster correlation of Y_{i0} (determines degree of common support)
    - we use **ICC = 0.1, 0.2, 0.3**
  - curvature parameter of metric of Y
    - allow observed Y metric to be a nonlinear (quadratic) transformation of the true Y metric
    - define curvature as ratio of derivative of transformation function at 95^{th} percentile to derivative at 5^{th} percentile
    - we use **curvature = 1/5, 1/3, 2/3, 1, 1.5, 3, 5**

simulation model

- we simulate observed data from 500 schools, each containing 500 students
  - large samples ensure precise estimation of school effects
  - assume Y is measured without error (rel.=1.0)
  - assume Y_0 is measured without error
  - assume ICC of gains for students with average initial score is 0.40 (based on extant research)

four VAM estimators

- model A: linear, no heterogeneity

$$Y_{ij} = \gamma_{10}^{A}\left(Y_{i0} - \bar{\bar{Y}}_0\right) + \Delta_j^A + e_{ij}^A$$

- model B: quadratic, no heterogeneity

$$Y_{ij} = \gamma_{10}^{B}\left(Y_{i0} - \bar{\bar{Y}}_0\right) + \gamma_{10}^{B}\left(Y_{i0} - \bar{\bar{Y}}_0\right)^2 + \Delta_j^B + e_{ij}^B$$

- model C: linear, with heterogeneity

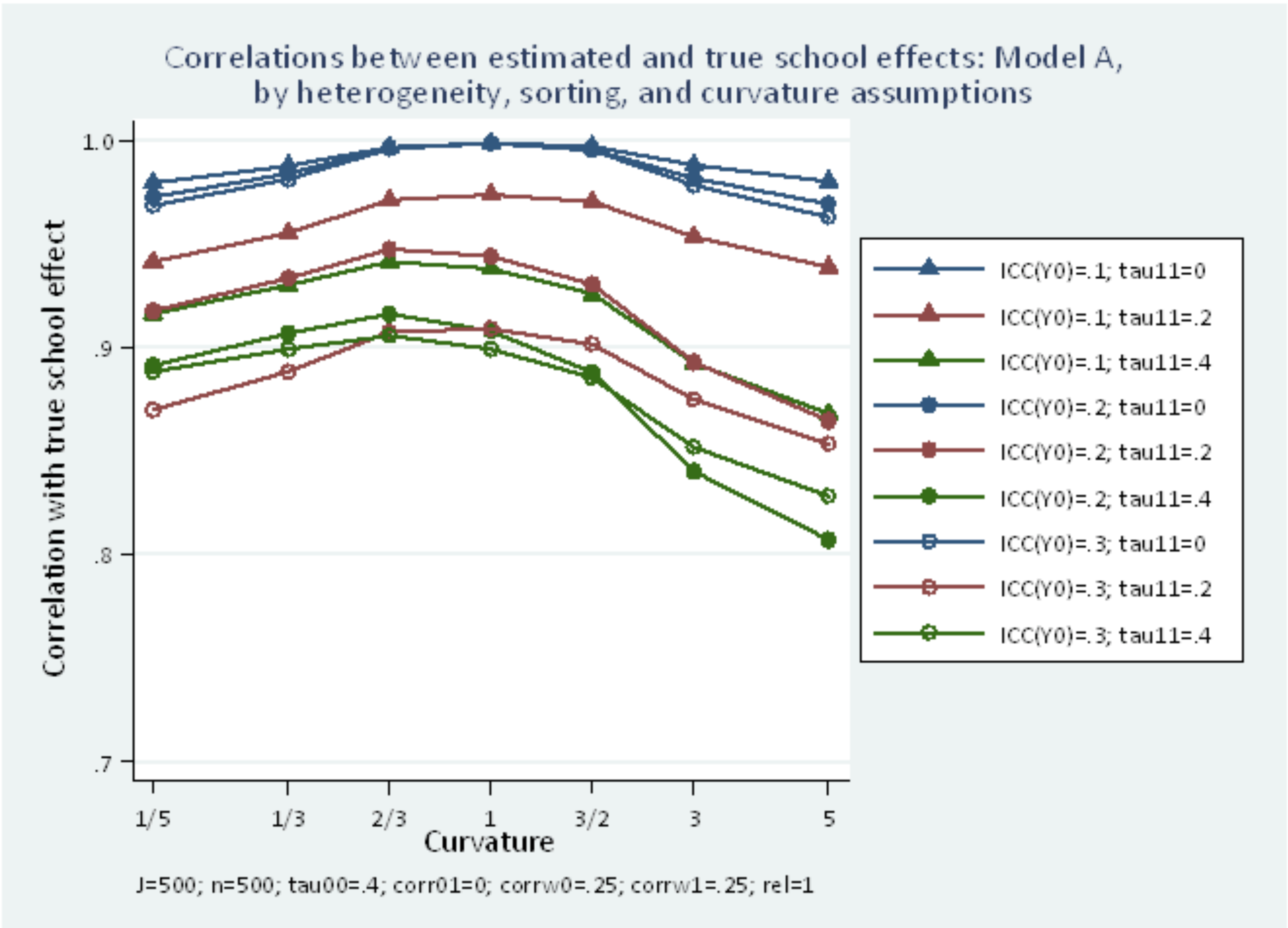
$$Y_{ij} = \left(\gamma_{10}^{C} + u_{1j}^{C}\right)\left(Y_{i0} - \bar{\bar{Y}}_0\right) + \Delta_j^C + e_{ij}^C$$

- model D: quadratic, with heterogeneity

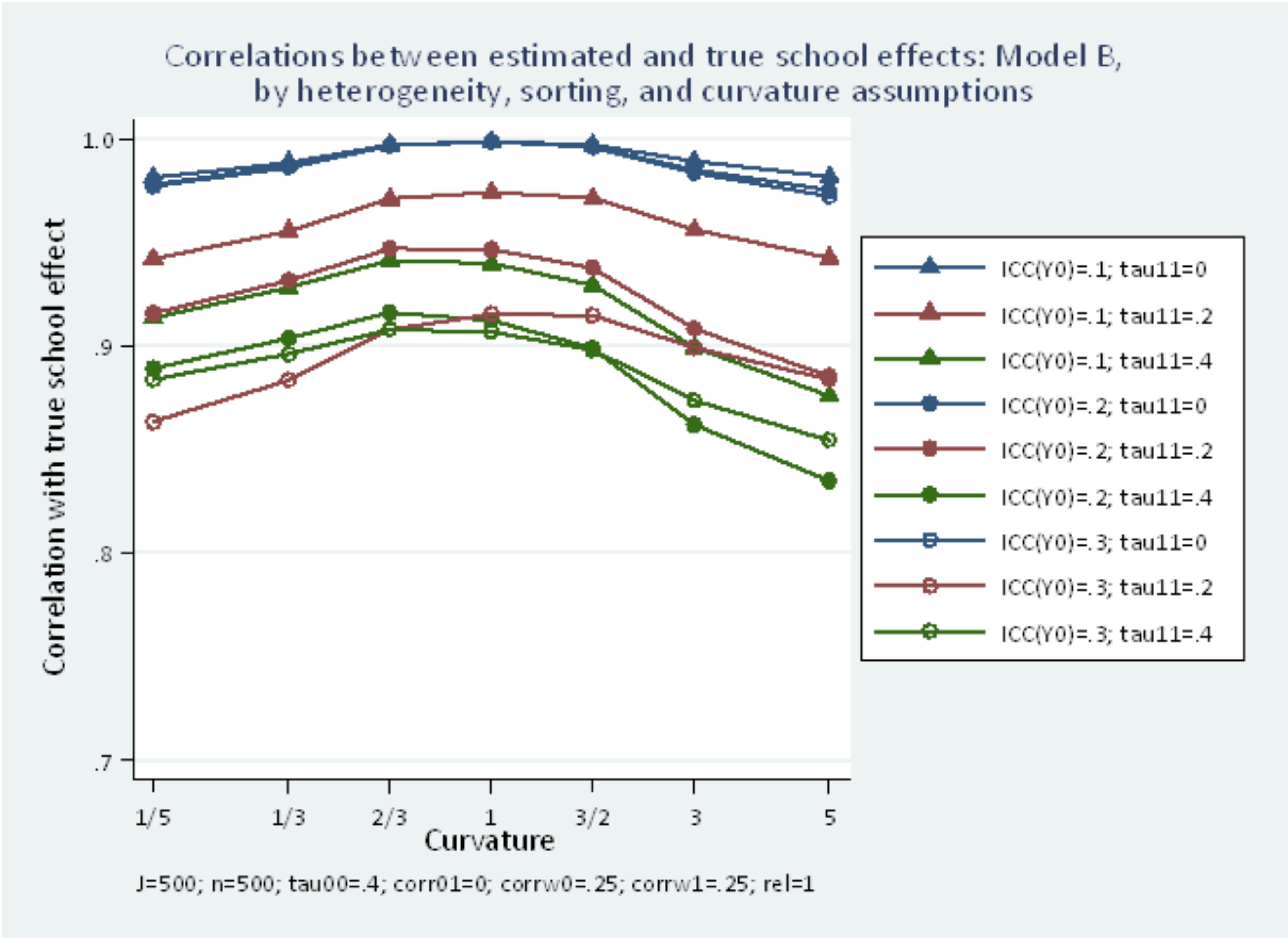
$$Y_{ij} = \left(\gamma_{10}^{D} + u_{1j}^{D}\right)\left(Y_{i0} - \bar{\bar{Y}}_0\right) + \left(\gamma_{20}^{D} + u_{2j}^{D}\right)\left(Y_{i0} - \bar{\bar{Y}}_0\right)^2 + \delta_j^D + e_{ij}^D$$

  - note: $\widehat{\Delta}_j^D = \hat{\delta}_j^D + \hat{u}_{2j}^D Var(Y_{i0})$

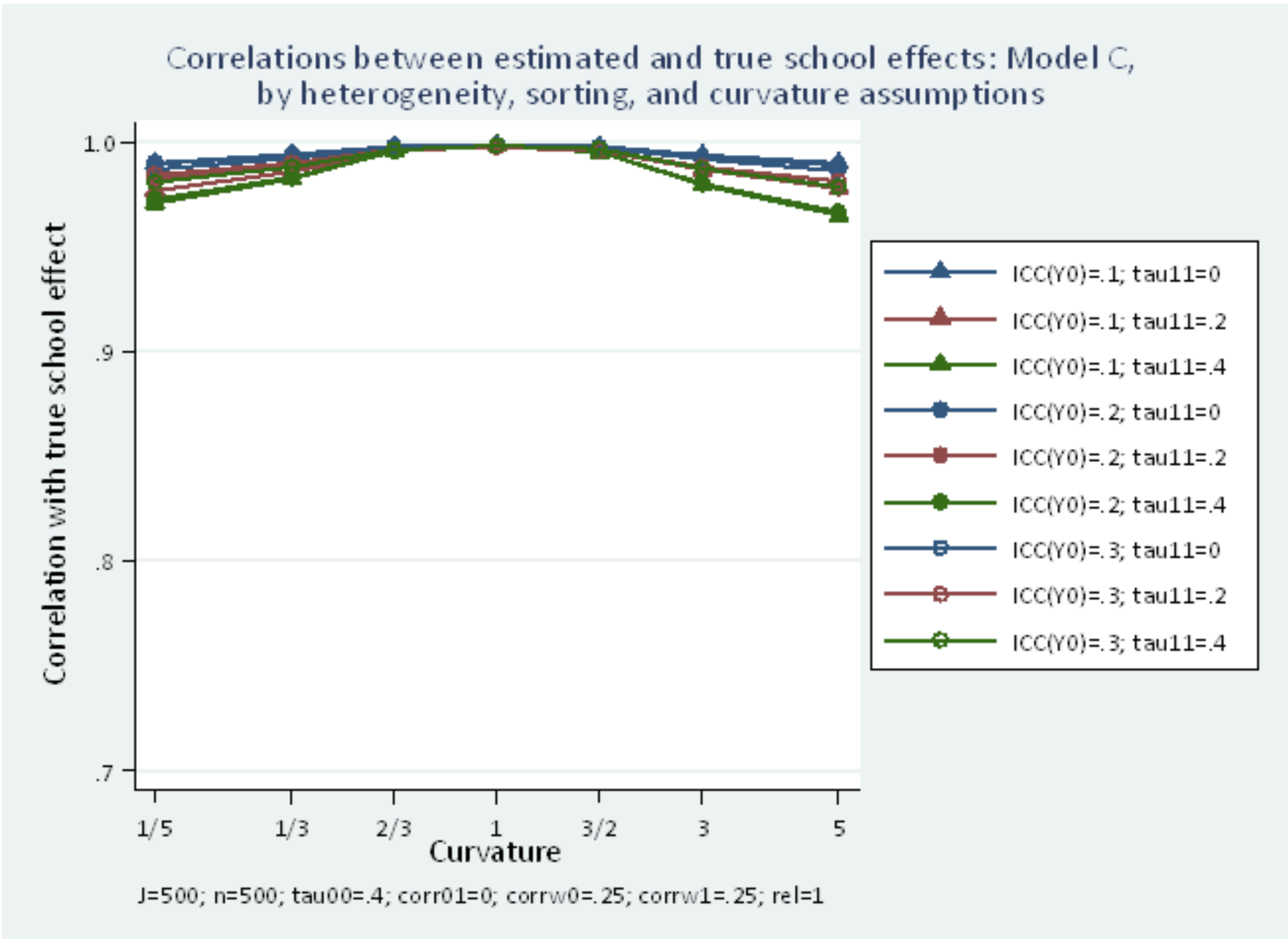
correlation of VAM A estimates and true school effects



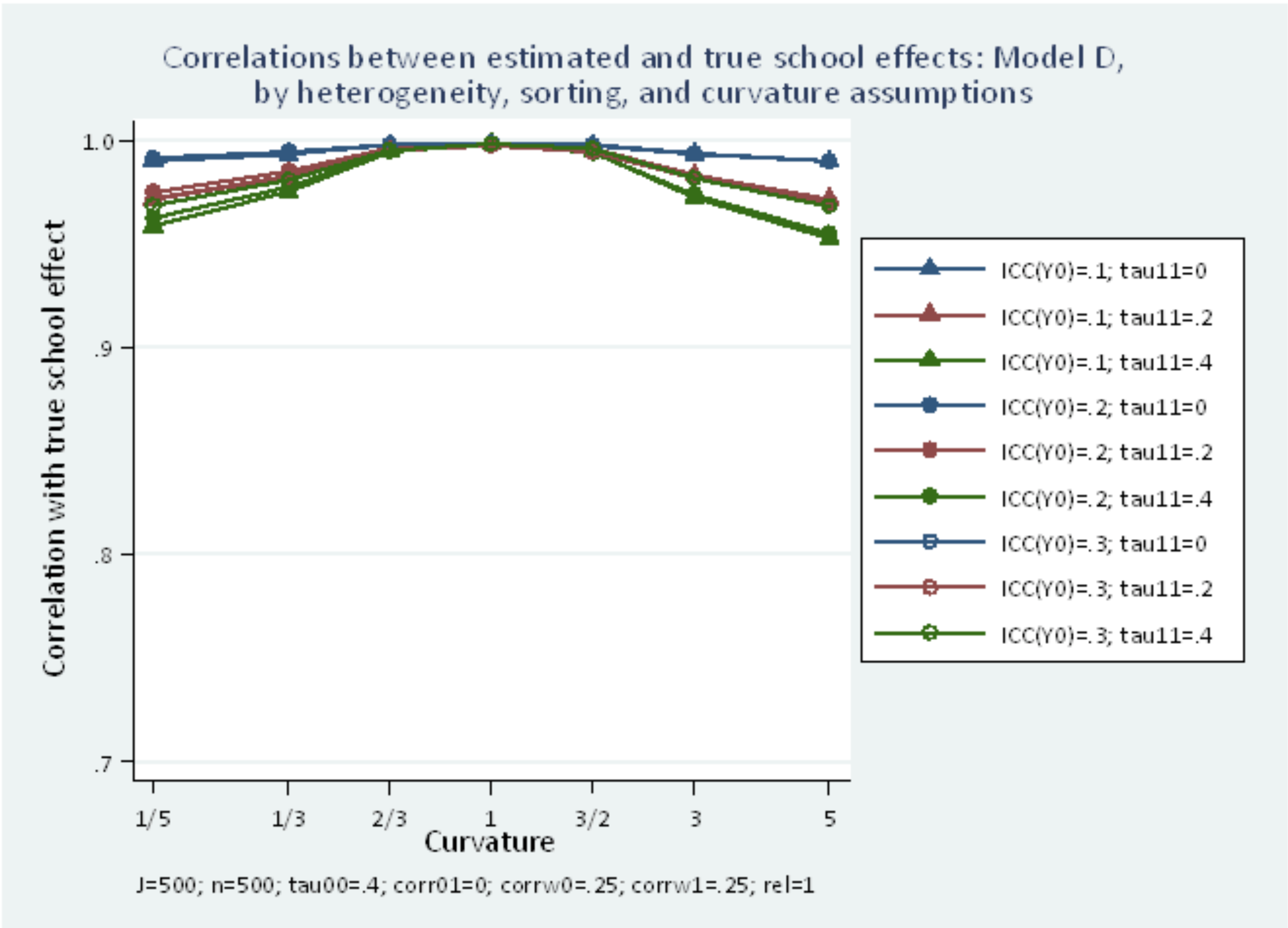
correlation of VAM B estimates and true school effects



correlation of VAM C estimates and true school effects



correlation of VAM D estimates and true school effects



conclusions

- sorting among schools need not degrade estimates of school effects, so long as homogeneity, interval metric (and ignorability) hold.
- heterogeneity of effects degrades estimates in the presence of sorting (lack of common support) and/or failure of interval metric assumption.
- failure of interval metric assumption degrades estimates, particularly in presence of heterogeneity of effects or incorrect specification of functional form.
- if there is heterogeneity of effects, models that allow for it perform substantially better than those that do not.

caveats

- many other potential sources of error in VAMs
  - bias due to violation of ignorability assumption
  - bias due to violation of SUTVA assumption
  - small school sizes may lead to substantially greater sampling variation
  - small school sizes may lead to less common support
  - measurement error in initial score Y_0
  - measurement error in outcome score Y
  - more complex potential outcome processes (our is exceedingly simplified)