Some Econometric Considerations for Value-Added Modeling

National Conference on Value-Added Modeling
April 24, 2008
Wisconsin Center for Education Research
Madison, WI

Jeff Wooldridge
Michigan State University

- Lots of recent research in evaluating different approaches to VAM, and the properties of different estimation approaches.
- Interest in testing underlying assumptions of the different approaches.
- How can recent (and not-so-recent) advances in panel data econometrics contribute to VAM?
- Issues at (say) the student level. (i) Unobserved student heterogeneity is likely to be correlated with observed inputs. (ii) Lagged dependent variable with serially correlated errors. (iii) Inputs arguably violate the strict exogeneity assumption, even after netting out unobserved heterogeneity. ("Feedback effects.")

“Static” Models

- Model in “standard” panel data notation:

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta}_t + c_i + u_{it},\ t = 1,\ldots,T, \tag{1}$$

where t can index grade and i indexes student. Here, y_{it} can be the achievement level, A_{it}, or the or change in achievement, ΔA_{it}.

- Well known how to obtain the “gain score” model from a more general VAM with cumulative effects.

- Especially in the level model, might be important to included lagged inputs in $\mathbf{x}_{it}$ (for example, class size).

- Issues in (1): (i) Large dimension for $\mathbf{x}_{it}$ (estimating teacher value added). (ii) Correlation between c_i and $\mathbf{x}_{it}$. (iii) Correlation between u_{it} and $\mathbf{x}_{ir}$ for any r. (Distributed lag dynamics? Contemporaneous endogeneity, that is, $Cov(\mathbf{x}_{it}, u_{it}) \neq \mathbf{0}$? Feedback, for example, $Cov(\mathbf{x}_{i,t+1}, u_{it}) \neq \mathbf{0}$?) (iv) Serial correlation in $\{u_{it} : t = 1, \ldots, T\}$. (v) Unbalanced panel: students drop out, enter sample. (vi) Does "large-N" asymptotics make sense, where N is the number of students? (Rothstein: estimating teacher effects.) (vii) Are the slopes truly constant across i? (Hierarchical Linear Models.) (viii) Does the impact of unobserved endowments change with grade? $c_i \rightarrow \eta_t c_i$?

- Language: c_i is a "student unobserved effect" or "heterogeneity." If $\mathbf{x}_{it}$ includes school dummies, should we call these "school *fixed* effects"? (If students are changing schools, why are they? If teachers change schools, why?)
- Standard estimation methods: (i) Pooled OLS, which maintains $Cov(\mathbf{x}_{it}, c_i) = 0$ and $Cov(\mathbf{x}_{it}, u_{it}) = 0$ for all t. Need to make inference robust to serial correlation (and heteroskedasticity) (ii) Fixed Effects, which maintains $Cov(\mathbf{x}_{ir}, u_{it}) = 0$, all r and t (strict exogeneity conditional on c_i). Can make a strong case for making inference robust to general serial correlation/heteroskedasticity in $\{u_{it} : t = 1, \ldots, T\}$.

However, with complicated clustering structures, how do "cluster-robust" standard errors behave? Hansen (2007, *Journal of Econometrics*) shows they can work with similar group sizes and groups, but what about large dimension of $\mathbf{x}_{it}$?

(iii) First Differencing, which (essentially) uses same assumption as FE. Again, robust inference. Less attractive than FE with general missing data patterns.

(iv) Random Effects: Strict exogeneity plus $Cov(\mathbf{x}_{it}, c_i) = 0$. Can and should make inference robust to general second moment structure (generalized estimating equations). (v) Correlated random effects: Write $c_i = \psi + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i$ (Mundlak, 1978) and use POLS or RE on

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + a_i + u_{it},\ t = 1,\ldots,T. \tag{2}$$

- Estimate of $\boldsymbol{\beta}$ is the FE estimator. (In this formulation, $\mathbf{x}_{it}$ can include interactions with time dummies.)

● Suppose we use the less restrictive Chamberlain (1982) device, $c_i = \psi + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \ldots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + a_i$, and estimate

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta}_t + \mathbf{x}_{i1}\boldsymbol{\lambda}_1 + \ldots + \mathbf{x}_{iT}\boldsymbol{\lambda}_T + a_i + u_{it},\ t = 1, \ldots, T \qquad (3)$$

by POLS or random effects. Still identical to the FE estimates for $\boldsymbol{\beta}_t$.

● Chamberlain does not impose a functional form; it is a linear projection. Can get more efficient estimation by minimum distance.

- FE applies more easily to unbalanced panels.
- Can add time-constant variables to (2) or (3) – say, family background.
- Aside: Can also apply Mundlak or Chamberlain to nonlinear models (say, if y_{it} is a fraction), and they often yield similar estimates of partial effects even if pattern of $\hat{\lambda}_t$ is far from constant.

- Under strict exogeneity,

$$E(u_{it}|c_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = 0,\ t = 1, \ldots, T, \tag{4}$$

a simple test of the "random effects" assumption, $Cov(\mathbf{x}_{it}, c_i) = \mathbf{0}$ is $H_0 : \boldsymbol{\xi} = \mathbf{0}$. Can make this test fully robust using "cluster" option. With RE, should include as many time-constant controls as possible.

- Pitfalls with directly comparing FE and RE: (i) Difficult to make robust (usual Hausman test maintains efficiency of RE under H_0). (ii) Sometimes cohort/grade effects are included in the comparison, leading to wrong degrees-of-freedom.

• Before testing for correlated random effects, makes sense to test for feedback. Construct the (artificial) model

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{i,t+1}\boldsymbol{\delta} + c_i + e_{it},\ t = 1,\ldots,T-1, \tag{5}$$

where $\mathbf{w}_{it}$ is a subset of $\mathbf{x}_{it}$, and estimate it by FE, testing $H_0 : \boldsymbol{\delta} = \mathbf{0}$. Note that contemporaneous variables are generally included along with the lead values.

• Todd and Wolpin (2003): first differences. Rothstein (2007): Chamberlain's correlated random effects approach.

Extensions Assuming Strict Exogeneity

- Incorporate time-varying effects into $\mathbf{x}_{it}$ and allow for individual-specific slopes:

$$y_{it} = \alpha_t + \mathbf{x}_{it}\mathbf{b}_i + c_i + u_{it}. \tag{6}$$

Use usual FE, that is, act as if $\mathbf{b}_i = \boldsymbol{\beta}$ for all i. Let $\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$. Can show under usual strict exogeneity assumption that $\hat{\boldsymbol{\beta}}_{FE}$ is consistent for the population average effect if

$$E(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = E(\mathbf{b}_i) \tag{7}$$

- Might want to take out more individual heterogeneity in $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$, for example, use a "random trend approach": For each i, regress $\mathbf{x}_{it}$ on $1, t$ across time (if $T \geq 3$). If $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{g}_i t + \mathbf{r}_{it}$, $\mathbf{b}_i$ can be arbitrarily correlated with $(\mathbf{f}_i, \mathbf{g}_i)$.
- Model $\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\Pi}(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})' + \mathbf{d}_i$ where $\mathbf{h}_i = (\bar{\mathbf{x}}_i, \mathbf{z}_i)$ and $E(\mathbf{d}_i | \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{z}_i) = \mathbf{0}$ (Raudenbush and Bryk, 2002).For example, $\mathbf{z}_i$ might include a baseline test score and $\mathbf{x}_{it}$ includes school or teacher indicators. Substitute:

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta} + [(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}}) \otimes \mathbf{x}_{it}]\boldsymbol{\pi} + c_i + \mathbf{x}_{it}\mathbf{d}_i + u_{it} \tag{8}$$

Remove the additive effect, c_i, using the within transformation:

$$\ddot{y}_{it} = \ddot{\alpha}_t + \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + [(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}}) \otimes (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)]\boldsymbol{\pi} + \ddot{\mathbf{x}}_{it}\mathbf{d}_i + \ddot{u}_{it}. \quad (9)$$

- The interactions might be significant even though the estimates of $\boldsymbol{\beta}$ (population averaged effect) might be similar to basic FE.
- Or, just add $\mathbf{h}_i$ and $(\mathbf{h}_i - \bar{\mathbf{h}}) \otimes \bar{\mathbf{x}}_i$ to the untransformed equation and use pooled OLS.
- Can use a generalized least squares method to account for the conditional heteroskedasticity in the variance-covariance matrix. But still use fully robust inference.
- Model for $E(\mathbf{b}_i|\mathbf{x}_i, \mathbf{z}_i)$ needs to be taken seriously.

- Grade-specific factor loads:

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta} + \eta_t c_i + u_{it}. \tag{10}$$

- Easy to see that usual FE consistently estimates $\boldsymbol{\beta}$ if $E(\ddot{\mathbf{x}}_{it}' c_i) = \mathbf{0}$.

Can estimate the η_t along with $\boldsymbol{\beta}$ (Ahn, Lee, and Schmidt, 2001, *Journal of Econometrics*). Might reject $\eta_t = 1, t = 2, \ldots, T$, with little effect on $\boldsymbol{\beta}$.

Shrinkage Estimation of VAM

- Shrinkage estimators of VA are usually motivated as empirical Bayes or best linear unbiased predictor. Can also derive it from predicting future outcome for a student assigned to the cluster. Let c index classroom, i index student. Standard two-level model with additive classroom effect:

$$y_{ci} = \mathbf{x}_{ci}\boldsymbol{\beta} + h_c + u_{ci} \equiv \mathbf{x}_{ci}\boldsymbol{\beta} + v_{ci},\ i = 1,\ldots,N_c, \tag{11}$$

and make the RE assumption that (h_c, u_{ci}) is independent of $\mathbf{X}_c$ $(N_c \times K)$, h_c is uncorrelated with u_{ci}, $\{u_{ci} : i = 1,\ldots,N_c\}$ are pairwise uncorrelated, and the variances are σ_h^2 and σ_u^2.

Now consider predicting the score (or gain score) for a new student in the classroom, y_{c,N_c+1}.

Let $\bar{v}_c = N_c^{-1}\sum_{i=1}^{N_c} v_{ci}$. Can show

$$
\begin{aligned}
E(y_{c,N_c+1}|\mathbf{X}_c, \mathbf{x}_{c,N_c+1}, y_{c1}, \ldots, y_{c,N_c}) &= \mathbf{x}_{c,N_c+1}\boldsymbol{\beta} + [\sigma_h^2/(\sigma_h^2 + \sigma_u^2/N_c)]\bar{v}_c && (12) \\
&\equiv \mathbf{x}_{c,N_c+1}\boldsymbol{\beta} + VA_c. && (13)
\end{aligned}
$$

- Can estimate $\boldsymbol{\beta}$ by pooled OLS, RE, or FE.

$$\widehat{VA}_c = [\hat{\sigma}_h^2/(\hat{\sigma}_h^2 + \hat{\sigma}_u^2/N_c)] \cdot (\bar{y}_c - \bar{\mathbf{x}}_c\hat{\boldsymbol{\beta}}). \qquad (14)$$

- When introduce time and pool across t, shrinkage VA calculations change with serial correlation, time-dependent variances.

- Equation (13), in this simplified setting, can motivate the Kane and Staiger (2008) approach for evaluating VAM. Do not need to control for student-specific covariates because of random assignment. But with reasonable sample sizes, one *should* control for these factors to reduce the error variance. For example, with classroom averages,

$$\bar{y}_{jp} = \alpha_p + \bar{\mathbf{x}}_{jp}\boldsymbol{\beta} + \gamma VA_{jp} + e_{jp},\ j = 1, 2. \tag{15}$$

- How does lack of strict exogeneity affect shrinkage calculations?

- What if the slopes are teacher specific, $\mathbf{b}_c = \boldsymbol{\beta} + \mathbf{d}_c$?

$$y_{c,N_c+1} = \mathbf{x}_{c,N_c+1}\boldsymbol{\beta} + \mathbf{x}_{c,N_c+1}\mathbf{d}_c + u_{c,N_c+1} \tag{16}$$

and the forecasted teacher effect,

$$E(\mathbf{x}_{c,N_c+1}\mathbf{d}_c + u_{c,N_c+1}|\mathbf{X}_c, \mathbf{y}_c) \tag{17}$$

depends on $\mathbf{x}_{c,N_c+1}$. With heterogeneous slopes on student characteristics, this calculaton of a teacher’s VA depends on the class composition as reflected in observed covariates.

- “Multiple Comparisons with the Best” literature to study rankings of teachers or schools. Can find a set of possibly “best” teachers at a particular confidence level. [Frontier production functions and technical inefficiency. Horrace and Schmidt (2000, *Journal of Applied Econometrics*).]

What if Strict Exogeneity is Violated in "Static" Models?

- Rothstein (2007, 2008) provides evidence of feedback in teacher assignment.
- Provided contemporaneous exogeneity holds, that is, $Cov(\mathbf{x}_{it}, u_{it}) = \mathbf{0}$, the "asymptotic bias" in the FE estimator is on the order of $O(T^{-1})$ but effectively assumes number of students gets large for each teacher. Little comfort in most VA applications. Bias in FD estimator does not disappear as T grows. (Caveat: u_{it} must be "weakly dependent" for FE result.)

- Common solution: Difference and use lags for IVs. Perhaps reasonable in gain score equation – so lagged inputs do not matter – but predicting $\Delta\mathbf{x}_{it}$ based on $\mathbf{x}_{i,t-1}$ need not work well.
- Can add moment conditions suggested by Arellano and Bover (1995, *Journal of Econometrics*), assuming

$$Cov(\Delta\mathbf{x}_{it}, c_i) = 0. \tag{18}$$

Gives moment conditions in differences and levels:

$$E[\mathbf{x}'_{i,t-1}(\Delta y_{it} - \Delta\alpha_t - \Delta\mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0},\ t = 2,\ldots,.T \quad (19)$$

$$E[\Delta\mathbf{x}'_{it}(y_{it} - \alpha_t - \mathbf{x}_{it}\boldsymbol{\beta})] = \mathbf{0},\ t = 2,\ldots,T. \quad (20)$$

- Does not seem useful for estimating teacher or school VA, but perhaps for estimating effects of inputs. (The change in class size is uncorrelated with heterogeneity, but initial class size can be, and future changes can depend on idiosyncratic shocks.)

Dynamic VAMs

- Well known how to get the equation

$$A_{it} = \lambda A_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + \psi_t c_i + r_{it} - \lambda r_{i,t-1},\ t = 1,2,\ldots \tag{21}$$

$$\equiv \lambda A_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + \psi_t c_i + u_{it} \tag{22}$$

from a cumulative effects model with constant decay – a particular "rational distributed lag model" – where r_{it} is the error in the original cumulative effects model.

- Generally, two sources of "endogeneity" in using pooled OLS: the unobserverd heterogeneity and serial correlation in $\{u_{it}\}$. (Serial correlation with lagged dependent variable makes sense here.)
- Still, $\psi_t c_i$ might be unimportant, and (22) often estimated without unobserved effect:

$$A_{it} = \lambda A_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}. \tag{23}$$

- If $\{r_{it} : t = 1, \ldots, T\}$ is pure measurement error, might make sense to assume it is serially uncorrelated. In practice, errors from DL models often have serial correlation.

- It is *possible* that $\{u_{it} = r_{it} - \lambda r_{i,t-1}\}$ is serially uncorrelated, and so OLS consistently estimates λ and $\boldsymbol{\beta}$. Can test for serial correlation after OLS estimation:

$$A_{it} \text{ on } A_{i,t-1}, \mathbf{x}_{it}, \hat{u}_{i,t-1},\ t = 2,\ldots,T \tag{24}$$

and use t test on $\hat{u}_{i,t-1}$.

- More powerful score test: model $r_{it} = \rho r_{i,t-1} + e_{it}$ and test $H_0 : \rho = \lambda$. Extend McClain and Wooldridge (1995, *Economics Letters*) for pure time series case. Set $\hat{r}_{i0} \equiv 0$ and build up $\hat{r}_{it} = \hat{\lambda}\hat{r}_{i,t-1} + \hat{u}_{it}$, $t = 1,\ldots,T$. Use $\hat{r}_{i,t-1}$ in place of $\hat{u}_{i,t-1}$ in (24).

- Extends to more flexible RDL models.
- When (23) holds when $\{u_{it}\}$ is serially uncorrelated, it has a nice feature: consistent estimation does not rely on strict exogeneity. Assignment is allowed to depend on the lagged test score and other observables.
- Rothstein (2008): using many lagged test scores might get closer to causal teacher effects. Does not estimate parameters of educational production function.

Sample Selection and Attrition

- Panels on students are often balanced because of missing data and attrition. FE has the advantage of allowing attrition (or entrance) to depend on student unobserved effect. RE requires attrition be uncorrelated with heterogeneity as well as idiosyncratic shocks.
- Simple test of attrition bias

$$y_{it} = \alpha_t + \mathbf{x}_{it}\boldsymbol{\beta} + \delta s_{i,t+1} + c_i + u_{it},\ t = 1, \ldots, T-1, \tag{25}$$

and test $H_0 : \delta = 0.$ Estimation by FE. Note that a correlated RE approach is more difficult to implement: can only project c_i onto time periods that are observed for student i.