SCM and fixed effects

Panel data are measured in both time and cross-sectional units. It is often useful to think in terms of an error components model for such data sets.

$$y_{it} = x_{it}\beta + u_{it}$$

where

$$u_{it} = \epsilon_i + \epsilon_t + \epsilon_{it}$$

Each cross-sectional unit in i has some common unobserved features ϵ_i and each time unit t has some common unobserved features ϵ_t. The remaining time and cross-sectional variation in outcome is reflected in ϵ_{it}. We treat the three components as mutually independent, as is typical.

Fixed time and cross-sectional effects are accommodated via dummy variable (D) regression.

$$y_{it} = x_{it}\beta + D\gamma + \epsilon_{it}$$

This is often referred to as the within groups effect model. The within groups model interpretation is more readily understood by double residual regression. First project X and Y onto D and keep the residuals (left nullspace component) from each.

$$P_D X = D\left(D^T D\right)^{-1} D^T X$$

$$P_D Y = D\left(D^T D\right)^{-1} D^T Y$$

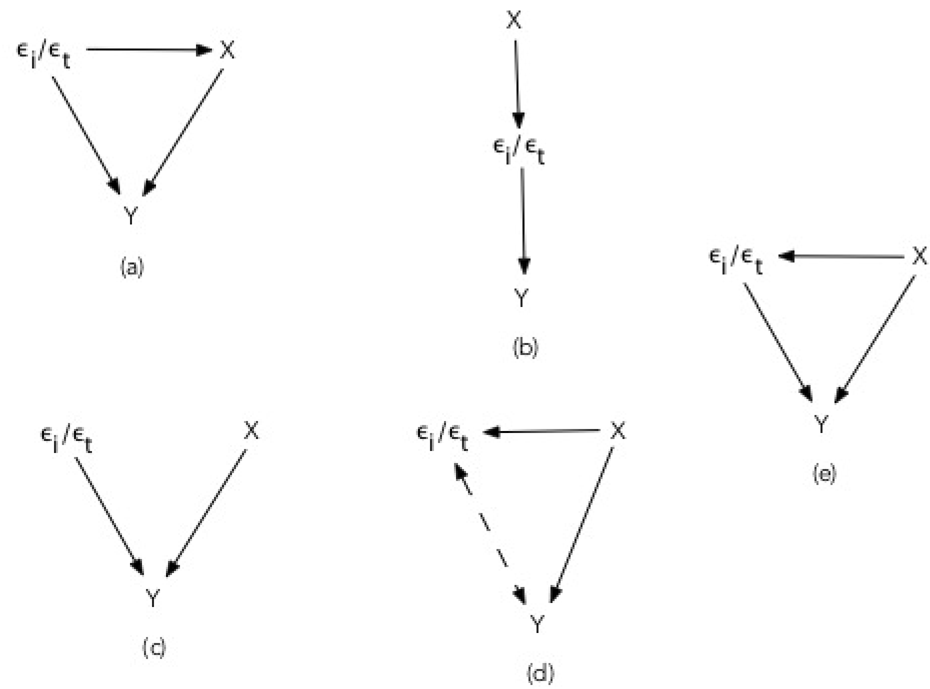
$$e_x = X - P_D X = \left(I - P_D\right) X = M_D X$$

$$e_y = M_D Y$$

where $M_D = I - P_D$ is the projection into the left nullspace of D. The residuals, e_x and e_y, are the deviations in X and Y, respectively, from the group means (within group variation). In the second step β is estimated by a regression of the residuals from Y, e_y, onto the residuals for X, e_x. Inclusion of D implies X, presumably the causal variable(s) of interest, is effectively $M_D X$ (X less group means) and consequently β reflects within group effects.

Consider the following simple DAGs (a graphical rendering of the mechanisms underpinning the causal relations amongst the variables) to evaluate the appropriateness of fixed effects in structural causal models (SCM).

$$y_{it} = x_{it}\beta + u_{it},\ u_{it} = \epsilon_i + \epsilon_t + \epsilon_{it}$$



DAG (a) is a classic fixed effects frame. The causal effect of $X \to Y$ is confounded by unobserved cross-sectional and/or time effects. Inclusion of fixed effects provides the appropriate adjustment for identification. DAG (c) does not require fixed effects adjustment for identification but fixed effects reduce residual variation in Y without causing harm.

In DAG (b), X is ancestor to the unobserved cross-section and/or time effects and consequently X is blocked from Y by the inclusion of such effects. The causal effect is not identified by fixed effects and in practice, as X and D are highly colinear, the estimation of β will be wildly unstable.

DAG (d) might be mistaken for an error components process — the "error components" do not map into outcome.[1] The "error components" in DAG (d) are colliders. Exclusion of D leaves the causal effect $X \to Y$ unconfounded but inclusion of fixed effects is confounding.

[1] Suppose an additional path from the error components to outcome describes the data generating process, fixed effects still fail to identify the causal effect of X on Y. That is, there is a back-door path into X if we adjust by fixed effects or if we don't adjust by the error components. Apparently, this variation on DAG (d) calls for instrumental variable identification of the causal effect $X \to Y$.

DAG (e) represents a within-between group effect frame. The total effect of $X \to Y$ simply involves $E[Y \mid X = x]$ (applying do-calculus rule 2, action/observation interchange) but can be decomposed into within group and between group effects. As in DAG (b) the path through the error components is the between group effect and the direct path is the within group effect (conditional on the error components). The within-between group effect design is

$$E\left[Y \mid X = x\right] = \beta_w \left(x - \bar{x}_j\right) + \beta_b \overline{x}_j = \beta_w M_D x + \beta_b P_D x$$

How to distinguish these frames?

As always, utilize background knowledge to frame the problem. Test whether X and Y are independent (which is not the case in any of the above DAGs). Error components are independent of X only in DAG (c). X and Y are independent conditional on D only in DAG (b).

Differentiating between DAGs (a), (d), and (e) is more challenging. DAGs (a) and (e) both call for fixed effects. However, the total effect and within group effect are the same for DAG (a) while the within group effect in DAG (e) is only part of the total effect. Distinguishing DAGs (a) or (e) from DAG (d) is aided by more data. Suppose we observe another set of variables W ancestor to X, then we are able to distinguish DAG (a) or (e) from DAG (d). In DAG (a) or (e), X is a collider between the error components and W and Y is a collider, hence, W is independent of D. On the other hand, W and D are only independent conditional on X in DAG (d) while conditioning on X makes W and D dependent in DAG (a) or (e). Also, W is an instrument aiding identification in DAG (d).