SCM, random coefficients, fixed effects, and clustering

Abadie, et al [2017] describe clustering as a design issue. Clustering is a *sampling design* issue if sampling follows a two-stage process. During the first stage, a sample of clusters is selected from a larger population of clusters. During the second stage, units are randomly sampled from the sampled clusters. Abadie, et al [2017] argue clustering in economics is more likely an *experimental design* issue. Clustering is an experimental design issue when clusters of units, rather than units, are assigned to a treatment so that the clusters or groups are correlated with treatment assignments.

Abadie et al [2017] indicate clustered standard errors are called for in a fixed effects regression when fixed effects are applied at the cluster or group level and there is heterogeneity in the treatment effects. Random coefficients associated with treatment is such a heterogeneous setting and we use random coefficients to illustrate when there is demand for clustered standard errors.

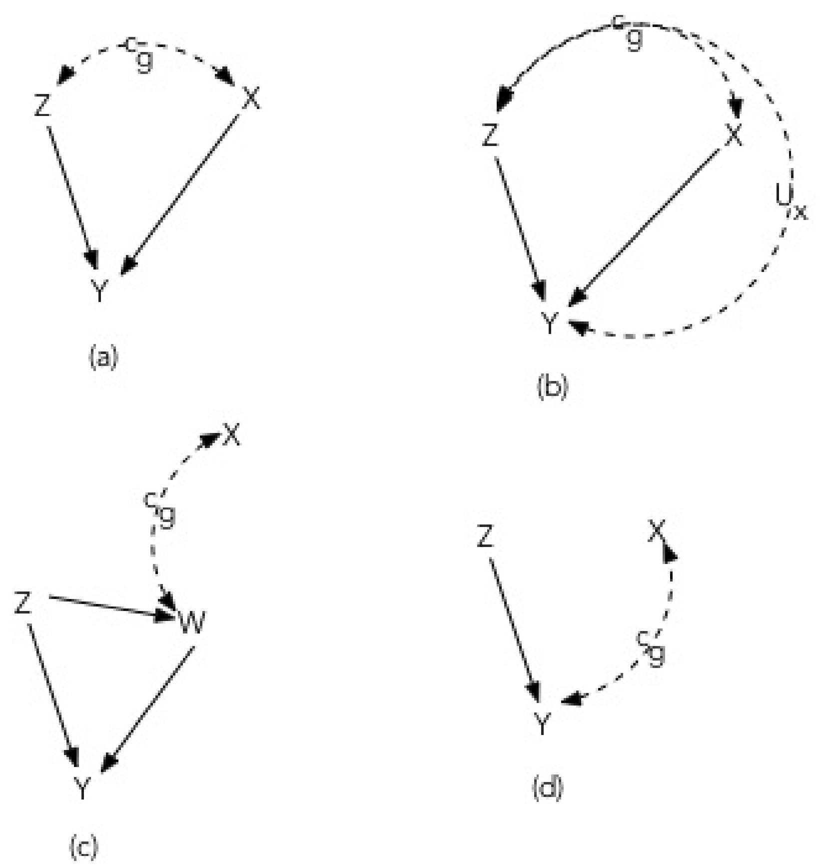
We utilize the experimental design clustering, random coefficients and fixed effects framework developed in Wooldridge [2011] along with DAGs (Pearl [2016]) to illustrate their structural causal modeling (SCM) implications. Consider the following random coefficients linear model in which each group or cluster g contains $m = 1, \ldots, M_g$ units

$$y_{gm} = \alpha + x_g\beta + z_{gm}\gamma_g + \nu_{gm} \tag{1}$$

where x_g is a $1 \times K$ vector of covariates that vary only at the cluster (between-group) level and z_{gm} is a $1 \times L$ vector of covariates that vary within clusters as well as perhaps between groups. β is the group level effect and γ_g is the unit level effect which varies across groups or clusters g . Our emphasis is primarily the unit level effect (the random coefficient γ_g). Further, let the error contain an unobserved common group effect c_g and an idiosyncratic component u_{gm} .

$$\nu_{gm} = c_g + u_{gm} \tag{2}$$

Next, we consider implications of this model in various causal frames depicted by DAGs (directed acyclic graphs) where the focus is the causal effect of $Z \rightarrow Y$. The graphs make the role of the unobserved common group component, c_g , explicit.



DAG (a) indicates the case in which adjusting for the group covariates X addresses confounding of the causal effect(s) of interest. DAG (b) is similar to (a) except that some subset of the group covariates X are unmeasured or unobserved. Therefore, unlike DAG (a) adjusting by X is insufficient to address confounding. Nonetheless, DAGs (a) and (b) can be readily accommodated via fixed effects as fixed effects adjust the confounding common or between-group component, c_g . Further, the effects of interest are within-group effects.

DAGs (c) and (d) are atypical fixed effects DGPs as there is no demand for fixed effects or c_g adjustment. Nonetheless, for a more complete survey we examine implications of cluster-based fixed effects for these DGPs. DAG (c) introduces W as a descendant of Z and c_g . As a result W embodies both within- and between-group effects. DAGs (c) and (d) require no adjustment of X , c_g (fixed effects), or W to identify the total effect of Z on Y . Adjustment by W identifies the group effect β while fixed effects regression suppresses the group effect. In DAG (c), adjustment by W identifies the direct effect of Z on Y while fixed effects identifies the within-group effect. In DAG (d), direct and total effects are the same.

Fixed effects regression

Now, focus on a fixed effects (within-group or cluster) regression where we identify the average effect γ .

$$y_{gm} - \overline{y}_g = (z_{gm} - \overline{z}_g)\,\gamma + \{(u_{gm} - \overline{u}_g) + (z_{gm} - \overline{z}_g)\,(\gamma_g - \gamma)\} \tag{3}$$

The term in braces is the error term. Even if u_g is randomly drawn from the same distribution, it is likely that the extra term $(z_{gm} - \overline{z}_g)\,(\gamma_g - \gamma)$ is correlated and/or heteroskedastic calling for a robust variance estimator. Clustered standard errors are drawn from a robust sandwich variance estimator for $\hat{\gamma}_{FE}$

$$\left(\sum_{g=1}^{G} Z_g^T Z_g\right)^{-1} \left(\sum_{g=1}^{G} Z_g^T \hat{u}_g \hat{u}_g^T Z_g\right) \left(\sum_{g=1}^{G} Z_g^T Z_g\right)^{-1} \tag{4}$$

where Z_g is the $M_g \times L$ matrix of within-group deviations and $\hat{u}_g$ is the $M_g \times 1$ vector of fixed effects residuals. The variance estimator is asymptotically consistent based on (the law of) large numbers (of clusters) G .

Rather than fixed effects regression, suppose we employ pooled-OLS regression with any appropriate adjustment necessary to identify the average effect γ (for DAG (a) inclusion of X for DAGs (c) and (d) no adjustment is needed). Clustered standard errors are again drawn from a sandwich variance estimator for $\hat{\gamma}$

$$\left(\sum_{g=1}^{G} W_g^T W_g\right)^{-1} \left(\sum_{g=1}^{G} W_g^T \hat{v}_g v_g^T W g\right) \left(\sum_{g=1}^{G} W_g^T W_g\right)^{-1} \tag{5}$$

where W_g is the $M_g \times (1 + K + L)$ matrix of all regressors (including a vector of ones for the intercept) for group g and $\hat{v}_g$ is the group g $M_g \times 1$ vector of pooled-OLS residuals.

Example: DAG (a)

As we emphasize the demand for clustered standard errors with fixed effects we suppress u_{gm} in the examples so that the only contributor to variation in the estimates is due to the heterogeneous term $(z_{gm} - \overline{z}_g)\,(\gamma_g - \gamma)$. The **DGP** is as follows. X is comprised of two covariates that vary strictly at the group or cluster level: x_{1g} varies from $1, \ldots, 100$ and x_{2g} varies from $50, \ldots, 1; 51, \ldots, 100$. That is, the number of groups or clusters is $G = 100$. z_{gm} varies with each cluster from $-5, \ldots, 4$ ($M_g = 10$ for each cluster) plus $x_1 + x_2$ (as Z and X are descendants of c_g). Finally, we center outcome at zero $y_{gm} = 0.5\,(x_{1g} - \overline{x}_1) + 1\,(x_{2g} - \overline{x}_2) + (z_{gm} - \overline{z}_g)\,\gamma_g$ where $\gamma_g = -30, \ldots, 69$ for $g = 1, \ldots, 100$. The average treatment effect of Z on Y is 19.5.

Linear regression of Y on Z adjusted by x_1 and x_2 effectively identifies the average total and direct effect. Also, Linear regression of Y on Z adjusted by cluster-based fixed effects identifies the average effect. On the other hand, omission of both X and cluster-based fixed effects (inclusion of only Z) produces substantial bias in the average causal effect, -0.19 rather than 19.5.

The error is due to the heterogeneous component $(z_{gm} - \overline{z}_g)(\gamma_g - \gamma)$ only. The estimated standard error for the fixed effect estimator of the average effect, $\hat{\gamma} = 19.4$, assuming a random sample of errors from the same population is 0.96. The Eicker-Huber-White heteroskedastic-consistent standard error is 1.22. On the other hand, the *cluster-adjusted standard error* is approximately three times larger than the standard estimator, 2.89.[1] The difference is entirely due to heterogeneity, if $\gamma_g = \gamma$ for all g the standard errors are the same. Abadie et al [2017] argue differences in standard errors is not sufficient to justify clustered standard errors. However, experimental design clustering combined with heterogeneous effects as in this setting creates demand for clustered standard errors.

Example: DAG (b)

The DGP is the same as that for DAG (a). Everything is the same except only x_1 is observed. Adjusting by x_1 is leaves the effect confounded. However, adjusting via fixed effects remains valid and fixed effects results are the same as for DAG (a) including demand for clustered standard errors. Omission of both x_2 and cluster-based fixed effects (inclusion of only x_1 and Z) produces substantial bias in the average causal effect, -0.55 rather than 19.5.

Fixed effect regression produces identical standard errors for the estimated average total effect as described for DAG (a). Inadequately adjusted (x_1 only) pooled-OLS regression clustered standard error is similar but smaller than that for fixed effect regression, 2.41 rather than 2.89.

Example: DAG (c)

For DAG (c) Z is the within-group component of that described for DAGs (a) and (b), $z_{gm} = -5, \ldots, 4$ for $z_{g1}, \ldots, z_{g10}$ in each cluster g. X is x_1 only,[2] and $w_{gm} = x_{1g} + 10z_{gm}$. Outcome is centered at zero $y_{gm} = 0.5\,(w_{gm} - \overline{w}) + (z_{gm} - \overline{z}_g)\,\gamma_g$.

The total effect of Z on Y is the direct effect plus the indirect or mediated (by W) effect, $19.5 + 10 * 0.5 = 24.5$. No adjustment is needed to identify the total effect, while adjustment by W identifies the direct effect. Fixed effect adjustment is not necessary but also identifies the total effect.

With sampling variation suppressed, fixed effect and pooled-OLS regression produce identical clustered standard errors for the estimated average total effect as described for DAG (a).

[1] With sampling variation suppressed, properly adjusted pooled-OLS regression cluster standard error is the same as that for fixed effect regression.

[2] x_2 is not included in the DGP as it is correlated with Z and inconsistent with DAGs (c) and (d).

Example: DAG (d)

The DGP is similar to that for DAG (c) except that W is eliminated leaving Z and x_1 independent ancestors to Y. Outcome is centered at zero $y_{gm} = 0.5\left(x_{1g} - \overline{x}_1\right) + \left(z_{gm} - \overline{z}_g\right)\gamma_g$. No adjustments are necessary to identify the average causal effect of Z on Y, γ. Nonetheless, fixed effects only mask the effect of x_1 leaving γ intact.

With sampling variation suppressed, fixed effect and pooled-OLS regression produce identical clustered standard errors for the estimated average total effect as described for DAG (a).