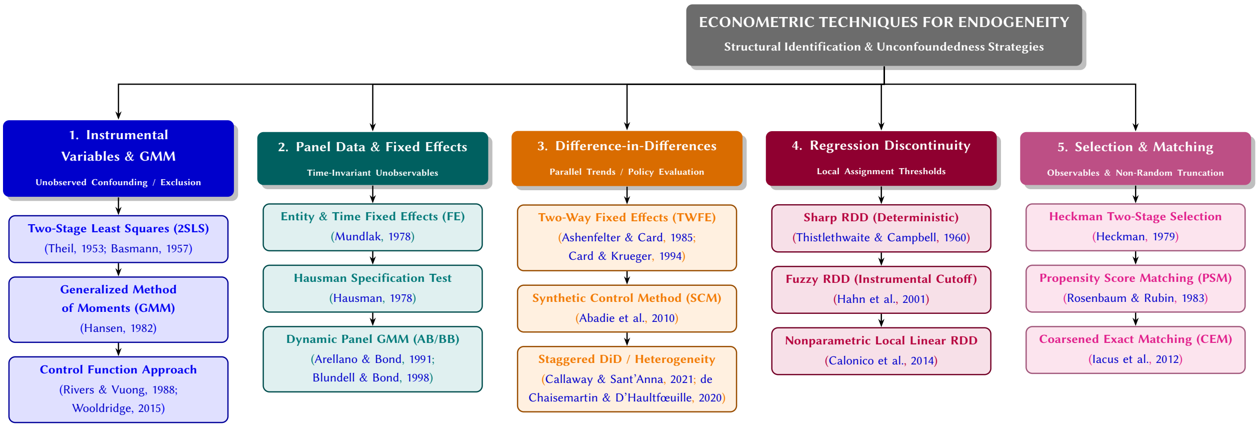
Econometrics **August 18, 2026**

Taxonomy of Causal Inference Methods

Instructor: Kalyan Kolukuluri *Content: Lecture Notes*

Note: *Notes may be distributed outside this class only with the permission of the Instructor. The author does not make claims on the authenticity of the information.*

1 Hierarchical Taxonomy of Econometric Techniques for Endogeneity



2 Identification Mechanics and Theoretical Principles

2.1 Instrumental Variables, 2SLS, and GMM Frameworks

Endogeneity occurs when an explanatory variable X is correlated with the error term u:

$$\mathrm{Cov}(X, u) \neq 0.$$

This correlation renders Ordinary Least Squares (OLS) biased and inconsistent. An instrumental variable Z isolates exogenous variation in X.

- **Relevance**:
  $$\mathrm{Cov}(Z, X) \neq 0.$$
  This condition is commonly assessed using the first-stage F-statistic.
- **Exogeneity or exclusion restriction**:
  $$\mathrm{Cov}(Z, u) = 0.$$
  This condition is generally justified using institutional knowledge and research design rather than established solely by statistical tests.

2.1.1 Two-Stage Least Squares

The two stages of 2SLS are

$$\begin{aligned} X_i &= \gamma_0 + \gamma_1 Z_i + v_i, && (1)\\ \widehat{X}_i &= \widehat{\gamma}_0 + \widehat{\gamma}_1 Z_i, && (2)\\ Y_i &= \beta_0 + \beta_1 \widehat{X}_i + u_i. && (3) \end{aligned}$$

2SLS is a special case of linear instrumental-variables GMM.

2.1.2 Generalized Method of Moments

GMM generalizes IV estimation through the moment restriction

$$E\left[Z_i\left(Y_i - X_i'\beta\right)\right] = 0. \tag{4}$$

With more instruments than coefficients, the model is overidentified and GMM combines the moment conditions using a weighting matrix.

2.1.3 Control Function Approach

The control-function approach estimates a first-stage residual and includes it in the structural equation. This explicitly controls for the endogenous component of the regressor.

2.2 Panel Data Econometrics and Fixed Effects

Panel data follow units $i = 1, \ldots, N$ over time periods $t = 1, \ldots, T$. A standard panel model is

$$Y_{it} = \beta X_{it} + \alpha_i + \lambda_t + \varepsilon_{it}, \tag{5}$$

where α_i represents unit-specific effects and λ_t represents time effects.

2.2.1 Fixed-Effects Estimator

The within transformation subtracts the unit-specific mean:

$$Y_{it} - \overline{Y}_i = \beta(X_{it} - \overline{X}_i) + (\varepsilon_{it} - \overline{\varepsilon}_i). \tag{6}$$

This removes the time-invariant effect α_i.

2.2.2 Hausman Specification Test

The Hausman test evaluates whether the random-effects estimator is consistent under

$$H_0 : \mathrm{Cov}(\alpha_i, X_{it}) = 0. \tag{7}$$

Rejection of the null generally supports the use of fixed effects rather than random effects.

2.2.3 Dynamic Panel GMM

Dynamic panel GMM addresses models containing lagged dependent variables:

$$Y_{it} = \alpha Y_{i,t-1} + \beta X_{it} + \mu_i + u_{it}. \tag{8}$$

Difference GMM and system GMM use lagged levels and differences as internal instruments under appropriate moment restrictions.

2.3 Difference-in-Differences and Event-Study Designs

Difference-in-Differences compares treated and control units before and after an intervention:

$$Y_{it} = \alpha + \beta\,\mathrm{Treat}_i + \gamma\,\mathrm{Post}_t + \delta_{DiD}(\mathrm{Treat}_i \times \mathrm{Post}_t) + \varepsilon_{it}. \tag{9}$$

The parameter δ_{DiD} is interpreted causally under the parallel-trends assumption.

2.3.1 Parallel Trends

In the absence of treatment, the average outcome change for the treatment group should equal the average outcome change for the control group.

2.3.2 Synthetic Control Method

The synthetic-control method constructs a weighted combination of untreated units to approximate the counterfactual outcome path of a treated unit.

2.3.3 Staggered Difference-in-Differences

When units adopt treatment at different times, conventional two-way fixed effects may be difficult to interpret under heterogeneous treatment effects. Group-time average treatment effects and event-study approaches can be used instead.

2.4 Regression Discontinuity Designs

Regression discontinuity designs exploit treatment assignment based on whether a running variable R_i crosses a cutoff c.

2.4.1 Sharp Regression Discontinuity

Treatment changes deterministically at the cutoff:

$$D_i = \mathbb{I}(R_i \geq c). \tag{10}$$

The local treatment effect is

$$\tau_{SRDD} = \lim_{r \downarrow c} E[Y_i \mid R_i = r] - \lim_{r \uparrow c} E[Y_i \mid R_i = r]. \tag{11}$$

2.4.2 Fuzzy Regression Discontinuity

In a fuzzy RDD, crossing the cutoff changes the probability of treatment but does not determine treatment perfectly. The cutoff indicator functions as an instrument for treatment.

2.4.3 Nonparametric Local Linear RDD

Local linear regression estimates the discontinuity using observations close to the cutoff and a data-driven bandwidth.

2.5 Selection Models and Matching Algorithms

2.5.1 Heckman Selection Model

The Heckman model corrects for non-random sample selection. A selection equation is first estimated, and the inverse Mills ratio is then included in the outcome equation.

2.5.2 Propensity Score Matching

The propensity score is

$$e(X_i) = P(D_i = 1 \mid X_i). \tag{12}$$

Matching treated and control units with similar propensity scores requires an unconfoundedness assumption conditional on observed covariates.

2.5.3 Coarsened Exact Matching

Coarsened exact matching temporarily groups continuous covariates into substantively meaningful intervals and then performs exact matching within the resulting strata.

3 Identification Assumptions and Research Design

The econometric method should follow the source of identification. Instrumental variables require credible relevance and exclusion restrictions. Fixed effects require sufficient within-unit variation and address time-invariant omitted variables. Difference-in-Differences requires a credible parallel-trends design. Regression discontinuity requires continuity of potential outcomes around the cutoff. Matching methods require selection on observables.

No estimator eliminates all forms of bias automatically. The credibility of an empirical result depends primarily on whether the assumptions generating the identifying variation are plausible in the institutional setting.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493–505.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2), 277–297.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of econometrics to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4), 648–660.

Basmann, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*, 25(1), 77–83.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115–143.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.

Card, D., & Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772–793.

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification of treatment effects by regression discontinuity designs. *Econometrica*, 69(1), 201–209.

Hansen, L. P. (1982). Large sample properties of Generalized Method of Moments estimators. *Econometrica*, 50(4), 1029–1054.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251–1271.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1–24.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1), 69–85.

Rivers, D., & Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3), 347–366.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

Theil, H. (1953). *Repeated Least Squares Applied to Economic Relations*. Central Planning Bureau, The Hague.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post-facto experiment. *Journal of Educational Psychology*, 51(6), 309–317.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2), 420–445.