cemmap
centre for microdata methods and practice

Heterogeneous coefficients, control variables, and identification of treatment effects

Whitney K. Newey
Sami Stouli

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP45/20

UKRI Economic and Social Research Council

HETEROGENEOUS COEFFICIENTS, CONTROL VARIABLES, AND IDENTIFICATION OF TREATMENT EFFECTS

WHITNEY K. NEWEY[†] AND SAMI STOULI[§]

Abstract. Multidimensional heterogeneity and endogeneity are important features of models with multiple treatments. We consider a heterogeneous coefficients model where the outcome is a linear combination of dummy treatment variables, with each variable representing a different kind of treatment. We use control variables to give necessary and sufficient conditions for identification of average treatment effects. With mutually exclusive treatments we find that, provided the generalized propensity scores (Imbens, 2000) are bounded away from zero with probability one, a simple identification condition is that their sum be bounded away from one with probability one. These results generalize the classical identification result of Rosenbaum and Rubin (1983) for binary treatments.

Keywords: Treatment effect; Multiple treatments; Heterogeneous coefficients; Control variable; Identification; Conditional nonsingularity; Propensity score.

1. Introduction

Models that allow for multiple treatments are important for program evaluation and the estimation of treatment effects (Cattaneo 2010; Heckman, Ichimura, Smith, and Todd 1998; Imai and van Dyk 2004; Imbens 2000; Graham and Pinto 2018; Lechner 2001; Wooldridge 2004). A general class is heterogeneous coefficient models where the outcome is a linear combination of dummy treatment variables and unobserved heterogeneity. These models allow for multiple treatment regimes, with each variable representing a different kind of treatment. These models also feature multidimensional heterogeneity, with the dimension of unobserved heterogeneity being determined by the number of treatment regimes.

Endogeneity is often a problem in these models because we are interested in the effect of treatment variables on an outcome, and the treatment variables are correlated with heterogeneity. Control variables provide an important means of controlling for

[†] Department of Economics, MIT, wnewey@mit.edu.
§ Department of Economics, University of Bristol, s.stouli@bristol.ac.uk.

endogeneity with multidimensional heterogeneity. For treatment effects, a control variable is an observed variable that makes heterogeneity and treatment variables independent when it is conditioned on (Rosenbaum and Rubin, 1983).

We use control variables to give necessary and sufficient conditions for identification of average treatment effects based on conditional nonsingularity of the second moment matrix of the vector of dummy treatment variables given the controls. This result is familiar in the binary treatment case, but its generalization to multiple treatments appears to be new. With mutually exclusive treatments we find that, provided the generalized propensity scores (Imbens, 2000) are bounded away from zero with probability one, a simple condition for identification is that their sum be bounded away from one with probability one. These results provide an important generalization of Rosenbaum and Rubin (1983)'s classical identification result for binary treatments.

2. Modeling of Treatment Effects

Let Y denote an outcome variable of interest, and X a vector of dummy variables $X(t)$, $t \in \mathcal{T} \equiv \{1, \ldots, T\}$, taking value one if treatment t occurs and zero otherwise, and ε a structural disturbance vector of finite dimension. We consider a heterogeneous coefficients model of the form

$$Y = p(X)'\varepsilon, \qquad p(X) = (1, X(1), \ldots, X(T))'. \tag{2.1}$$

This model is linear in the treatment dummy variables, with coefficients ε that need not be independent of X. We assume that the vector ε is mean independent of the endogenous treatments X, conditional on an observable control variable denoted V.

Assumption 1. For the model in (2.1), there exists a control variable V such that $E[\varepsilon|X, V] = E[\varepsilon|V]$.

The Rosenbaum and Rubin (1983) treatment effects model is included as a special case where $X \in \{0, 1\}$ is a treatment dummy variable that is equal to one if treatment occurs and equals zero without treatment, and

$$p(X) = (1, X)'.$$

In this case $\varepsilon = (\varepsilon_1, \varepsilon_2)'$ is two dimensional with ε_1 giving the outcome without treatment, and ε_2 being the treatment effect. Here the control variables in V would be observable variables such that Assumption 1 holds, i.e., the coefficients $(\varepsilon_1, \varepsilon_2)$ are

mean independent of treatment conditional on controls; this is the unconfoundedness assumption of Rosenbaum and Rubin (1983).

A central object of interest in model (2.1) is the average structural function given by $\mu(X) \equiv p(X)'E[\varepsilon]$; see Blundell and Powell (2003) and Wooldridge (2005). This function is also referred to as the dose-response function in the statistics literature (e.g., Imbens, 2000). When $X \in \{0, 1\}$ is a dummy variable for treatment, $\mu(0)$ gives the average outcome if every unit remained untreated and $\mu(1)$ the average outcome if every unit were treated, with $\mu(1) - \mu(0)$ being the average treatment effect. In general, the average effect of some treatment $t \in \mathcal{T}$ is

$$\mu(e_t) - \mu(0_T),$$

with $e_t = (0, \ldots, 0, 1, 0, \ldots, 0)'$ defined as a T-vector with all components equal to zero, except the tth, which is one, and 0_T a T-vector of zeros.

The conditional mean independence assumption and the form of the structural function $p(X)'\varepsilon$ in (2.1) together imply that the control regression function of Y given (X, V), $E[Y|X, V]$, is a linear combination of the treatment variables:

(2.2)
$$E[Y \mid X, V] = p(X)'E[\varepsilon \mid X, V] = p(X)'E[\varepsilon \mid V] = p(X)'q_0(V),\ q_0(V) \equiv E[\varepsilon \mid V].$$

The average structural function can thus be expressed as a known linear combination of $E[q_0(V)]$ from equation (2.2). By iterated expectations,

$$p(X)'E[q_0(V)] = p(X)'E[E[\varepsilon \mid V]] = \mu(X).$$

We use the varying coefficient structure of the control regression function (2.2) and the implied linear form of $\mu(X)$ to give conditions that are necessary as well as sufficient for identification.

3. Identification Analysis

A sufficient condition for identification of the average structural function is nonsingularity of the second moment matrix of the treatment dummies given the controls,

$$E[p(X)p(X)' \mid V],$$

with probability one. Under the maintained assumption that $E[p(X)p(X)']$ is nonsingular, this condition is also necessary.

Theorem 1 states our first main result. The proofs of all formal results are given in the Appendix.

Theorem 1. *Suppose that* $E[\|\varepsilon\|^2] < \infty$, $E[p(X)p(X)']$ *is nonsingular, and Assumption 1 holds. Then:* $E[p(X)p(X)'|V]$ *is nonsingular with probability one if, and only if,* $\mu(X)$ *is identified.*

When $X \in \{0, 1\}$ and $p(X) = (1, X)'$, the identification condition becomes the standard condition for the treatment effect model

$$Y = \varepsilon_1 + \varepsilon_2 X, \quad E[\varepsilon \mid X, V] = E[\varepsilon \mid V], \quad \varepsilon \equiv (\varepsilon_1, \varepsilon_2)'.$$

The identification condition is that the conditional second moment matrix of $(1, X)'$ given V is nonsingular with probability one, which is the same as

$$\text{var}(X \mid V) = P(V)[1 - P(V)] > 0, \quad P(V) \equiv \Pr[X = 1 \mid V], \tag{3.1}$$

with probability one, where $P(V)$ is the propensity score. Here we can see that the identification condition is the same as $0 < P(V) < 1$ with probability one, which is the standard identification condition.

With multiple treatments, because $p(X)$ includes an intercept, the identification condition is the same as nonsingularity of the variance matrix $\text{var}(X|V)$ with probability one. This result generalizes (3.1).

Theorem 2. $E[p(X)p(X)'|V]$ *is nonsingular with probability one if, and only if, the variance matrix* $var(X|V)$ *is nonsingular with probability one.*

Considerable simplification occurs with mutually exclusive treatments, which allows for the formulation of an equivalent condition for nonsingularity of $E[p(X)p(X)'|V]$ solely in terms of the generalized propensity scores (Imbens, 2000). This result generalizes the standard identification condition for binary X.

Theorem 3. *Suppose that* $\Pr[X(t) = 1|V] > 0$ *for each* $t \in \mathcal{T}$ *with probability one. With mutually exclusive treatments,* $E[p(X)p(X)'|V]$ *is nonsingular with probability one if, and only if,*

$$\Sigma_{s=1}^{T} \Pr[X(s) = 1 \mid V] < 1, \tag{3.2}$$

with probability one.

4. Discussion

The heterogeneous coefficients formulation we propose for multiple treatment effects reveals the central role of the conditional nonsingularity condition for identification. Because this condition is in principle testable, establishing that it is also necessary demonstrates testability of identification (e.g., Breusch, 1986). With mutually exclusive treatments, the formulation of the equivalent condition (3.2) thus relates testability of identification to the generalized propensity scores. This is a generalization of the relationship between testability of identification and the propensity score in the binary treatment case.

Conditions that are both necessary and sufficient are also important for the determination of minimal conditions for identification. In unpublished work Wooldridge (2004) considers a restricted version of our model with $E[X|\varepsilon, V] = E[X|V]$ and $E[p(X)p(X)'|\varepsilon, V] = E[p(X)p(X)'|V]$, and shows that $q_0(V)$ is identified if $E[p(X)p(X)'|V]$ is invertible. The additional conditional second moments assumption implies that his identification condition differs from ours. Thus his result and proof do not apply in our setting which only assumes conditional mean independence $E[\varepsilon|X, V] = E[\varepsilon|V]$, and our results show that conditional second moments independence is not necessary for identification in multiple treatment effect models. Graham and Pinto (2018) consider a related approach in work independent of the first version of this paper (Newey and Stouli, 2018) where the identification result (Lemma 1 in the Appendix) was derived. The conditional nonsingularity condition we propose is weaker than their identification condition, and we study necessity as well as sufficiency for identification of average treatment effects.

The identification results we obtain here are of general interest for the vast treatment effects literature (e.g., Imbens, 2004, Imbens and Wooldridge, 2009, and Athey and Imbens, 2017, for reviews) and complement existing results on identification of treatment effects.

Appendix A. Proofs

A.1. **Preliminary result.**

Lemma 1. *Suppose that* $E[\|\varepsilon\|^2] < \infty$ *and Assumption 1 holds. If* $E[p(X)p(X)'|V]$ *is nonsingular with probability one then* $q_0(V)$ *is identified.*

Proof. Let $\lambda_{\min}(V)$ denote the smallest eigenvalue of $E[p(X)p(X)'|V]$. Suppose that $\bar{q}(V) \neq q_0(V)$ with positive probability on a set $\widetilde{\mathcal{V}}$, and note that $\lambda_{\min}(V) > 0$ on $\mathcal{V}$ by assumption. Then

$$\begin{aligned} E\left[\left\{p(X)'\{\bar{q}(V) - q_0(V)\}\right\}^2\right] &= E\left[\{\bar{q}(V) - q_0(V)\}' E[p(X)p(X)' \mid V]\{\bar{q}(V) - q_0(V)\}\right] \\ &\geq E\left[\|\bar{q}(V) - q_0(V)\|^2 \lambda_{\min}(V)\right] \\ &\geq E\left[1(V \in \mathcal{V} \cap \widetilde{\mathcal{V}})\|\bar{q}(V) - q_0(V)\|^2 \lambda_{\min}(V)\right] \end{aligned}$$

By definition $\Pr[\widetilde{\mathcal{V}}] > 0$ and $\widetilde{\mathcal{V}} \subseteq \mathcal{V}$ so that $\widetilde{\mathcal{V}} \cap \mathcal{V} = \widetilde{\mathcal{V}}$. Thus the fact that $\|\bar{q}(V) - q_0(V)\|^2 \lambda_{\min}(V)$ is positive on $\widetilde{\mathcal{V}} \cap \mathcal{V}$ implies

$$E\left[1(V \in \mathcal{V} \cap \widetilde{\mathcal{V}})\|\bar{q}(V) - q_0(V)\|^2 \lambda_{\min}(V)\right] > 0.$$

We have shown that, for $\bar{q}(V) \neq q_0(V)$ with positive probability on a set $\widetilde{\mathcal{V}}$,

$$E\left[\left\{p(X)'\{\bar{q}(V) - q_0(V)\}\right\}^2\right] > 0,$$

which implies $p(X)'\bar{q}(V) \neq p(X)'q_0(V)$. Therefore, $q_0(V)$ is identified from $E[Y|X,V]$. □

A.2. **Proof of Theorem 1.** We first show that nonsingularity of $E[p(X)p(X)'|V]$ with probability one implies identification of $\mu(X)$. By Lemma 1, if $E[p(X)p(X)'|V]$ is nonsingular with probability one then $q_0(V)$ is identified, and hence $E[q_0(V)]$ also is. By $p(X)$ being a known function, $p(X)'E[q_0(V)] = \mu(X)$ is identified.

We now establish that nonsingularity of $E[p(X)p(X)'|V]$ with probability one is necessary for identification of $\mu(X)$. It suffices to show that singularity of $E[p(X)p(X)'|V]$ with positive probability implies that $\mu(X)$ is not identified, i.e., there exists an observationally equivalent $\overline{q}(V) \neq q_0(V)$ with positive probability such that $p(X)'E[\overline{q}(V)] \neq p(X)'E[q_0(V)]$ with positive probability. By nonsingularity of $E[p(X)p(X)']$ and linearity of $\mu(X)$, the conclusion holds if, and only if, there exists an observationally equivalent $\overline{q}(V) \neq q_0(V)$ with positive probability such that $E[\overline{q}(V)] \neq E[q_0(V)]$.

Suppose that $E[p(X)p(X)'|V]$ is singular with positive probability and let $\Delta(V)$ be such that $E[p(X)p(X)'|V]\Delta(V) = 0$. We have that $\Delta(V) \neq 0$ on a set $\widetilde{\mathcal{V}}$ with $\Pr[\widetilde{\mathcal{V}}] > 0$. For $J = T + 1$, define $\widetilde{\mathcal{V}}_j = \{v \in \widetilde{\mathcal{V}} : \Delta_j(v) \neq 0\}$, $j \in \{1, \ldots, J\}$. Then $\cup_{j=1}^J \widetilde{\mathcal{V}}_j = \{v \in \widetilde{\mathcal{V}} : \Delta(v) \neq 0\} = \widetilde{\mathcal{V}}$. Hence

$$0 < \Pr[\widetilde{\mathcal{V}}] = \Pr[\cup_{j=1}^J \widetilde{\mathcal{V}}_j] \leq \sum_{j=1}^J \Pr[\widetilde{\mathcal{V}}_j],$$

which implies that $\Pr[\widetilde{\mathcal{V}}_{j^*}] > 0$ for some $j^* \in \{1, \ldots, J\}$.

Set $\widetilde{\Delta}(v) = \Delta(v)$ for $v \in \widetilde{\mathcal{V}}_{j^*}$, and $\widetilde{\Delta}(v) = 0$ otherwise. By construction $\widetilde{\Delta}_{j^*}(V) \neq 0$, and letting

$$\widetilde{\widetilde{\Delta}}(V) = \text{sign}\{\widetilde{\Delta}_{j^*}(V)\}\frac{\widetilde{\Delta}(V)}{||\widetilde{\Delta}(V)||},$$

we have that $\widetilde{\widetilde{\Delta}}_{j^*}(V) > 0$ on $\widetilde{\mathcal{V}}_{j^*}$ and $||\widetilde{\widetilde{\Delta}}(V)|| = 1$, and hence $E[|\widetilde{\widetilde{\Delta}}(V)|] < \infty$ and $E[\widetilde{\widetilde{\Delta}}_{j^*}(V)] \neq 0$. Therefore $E[\widetilde{\widetilde{\Delta}}(V)] \neq 0$, which implies that $E[q_0(V) + \widetilde{\widetilde{\Delta}}(V)] \neq E[q_0(V)]$. The result follows. □

A.3. **Proof of Theorem 2.** The matrix $E[p(X)p(X)'|V]$ is of the form

$$E[p(X)p(X)' \mid V] = \begin{bmatrix} 1 & E[X' \mid V] \\ E[X \mid V] & E[XX' \mid V] \end{bmatrix}, \tag{A.1}$$

and is positive definite if, and only if, the Schur complement of 1 in (A.1) is positive definite (Boyd and Vandenberghe, 2004, Appendix A.5.5.), i.e., if, and only if,

$$E[XX' \mid V] - E[X \mid V]E[X' \mid V] = \text{var}(X \mid V),$$

is positive definite with probability one, as claimed. □

A.4. **Proof of Theorem 3.** For a vector $w \in \mathbb{R}^T$, let $\text{diag}(w)$ denote the $T \times T$ diagonal matrix with diagonal elements $w_1, \ldots, w_T$. For mutually exclusive treatments, the matrix $E[p(X)p(X)'|V]$ is of the form

$$E[p(X)p(X)' \mid V] = \begin{bmatrix} 1 & E[X' \mid V] \\ E[X \mid V] & \text{diag}(E[X \mid V]) \end{bmatrix}. \tag{A.2}$$

The matrix $\text{diag}(E[X|V])$ has diagonal elements $E[X(t)|V] = \Pr[X(t) = 1|V] > 0$, for each $t \in \mathcal{T}$, and hence is positive definite. Therefore, $E[p(X)p(X)'|V]$ is positive definite if, and only if, the Schur complement of $\text{diag}(E[X|V])$ in (A.2) is positive definite (Boyd and Vandenberghe, 2004, Appendix A.5.5.), i.e., if, and only if,

$$\begin{aligned} 0 < 1 - E[X' \mid V]\text{diag}(E[X \mid V])^{-1}E[X \mid V] &= 1 - \Sigma_{s=1}^{T} E[X(s) \mid V] \\ &= 1 - \Sigma_{s=1}^{T} \Pr[X(s) = 1 \mid V], \end{aligned}$$

with probability one, as claimed. □

References

Athey, S. and Imbens, G. (2017). The state of applied econometrics: causality and policy evaluation. *Journal of Economic Perspectives*. 31, 3–32.

Blundell, R., and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*. 36, 312–357.

Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Breusch, T. S. (1986). Hypothesis testing in unidentified models. *Review of Economic Studies*. 53, 635–651.

Cattaneo, M. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics.* 155, 138–154.

Graham, B. S. and Pinto, C. C. D. X. (2018) Semiparametrically efficient estimation of the average linear regression function. *eprint arXiv:1810.12511*.

Heckman, J. J., Ichimura, H., Smith, J. and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*. 66, 1017–1098.

Imai, K. and van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association.* 99, 854–866.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*. 87, 706–710.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity. *Review of Economics and Statistics*. 86, 4–29.

Imbens, G. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature.* 47, 5–86.

Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric evaluation of labour market policies in Europe*, 43–58. Physica, Heidelberg.

Newey, W. K. and Stouli, S. (2018). Heterogenous coefficients, discrete instruments, and identification of treatment effects. *eprint arXiv:1811.09837*.

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

Wooldridge, J. M. (2004). Estimating average partial effects under conditional moment independence assumptions. *Cemmap working paper* CWP03/04.

Wooldridge, J. M. (2005). Unobserved heterogeneity and the estimation of average partial effects. In *Identification and inference for econometric models*, pp.27-55.