ESTIMATING THE EFFECTS OF NATURAL EXPERIMENTS

Halbert White
University of California, San Diego

12 January 2005

Abstract This paper investigates methods for estimating the effects of natural experiments, especially those created by an intervention or structural change occurring at a specific point in time, such as a government policy intervention, a merger, or the formation or disintegration of a cartel. We draw on the extensive literature of treatment effects (Rubin, 1974; Rosenbaum and Rubin, 1983; Hahn, 1998; Heckman, Ichimura, and Todd, 1998; Hirano, Imbens, and Ridder, 2003), but the fact that the treatment constituted by the natural experiment may precede the measurement of relevant covariates requires us to develop and operate within a framework ensuring that the estimated effects of such treatments are not contaminated by confounding biases that can otherwise easily arise. We analyze both the common dummy variable approach to estimating the effects of interest and other more flexible methods related to the treatment effect estimators of Hahn (1998), and Hirano, Imbens, and Ridder (2003) (HIR). As we show, the dummy variable approach is valid only under very strong assumptions not plausible in applications. The Hahn and HIR estimators are more generally valid, but they can be computationally challenging and their properties are unknown for the time-series applications of interest here. We propose a new computationally convenient estimator for the effects of interest that shares many of the advantages of the Hahn and HIR estimators, but whose asymptotic properties can be straightforwardly analyzed for either time-series or cross-section applications. Because of its computational simplicity and known properties, this alternate quasi-nonparametric estimator should prove useful in applications. Our analytic framework has general utility in that it provides an explicit and extensive role for economic theory in identifying suitable and unsuitable covariates. We examine the attendant issues in detail, both from a theoretical perspective and from an empirical perspective.

Acknowledgement: The author is grateful for the comments and suggestions of the editor, two anonymous referees, Douglas Bernheim, Xiaohong Chen, Clive Granger, David Hendry, Paul Johnson, Jack Porter, Chris Stomberg, Timo Terasvirta, participants of the San Diego Predictive Methodology and Application in Economics and Finance Conference in Honor of Clive W.J. Granger, and participants in my seminars at the Stockholm School of Economics in May and December, 2003. All errors and omissions are the author's responsibility.

1 Introduction

Causality has been a long-standing interest of Clive Granger. His seminal paper setting forth the concept now known as "Granger Causality" (Granger, 1969) has spawned a vast literature. As Clive readily points out, Granger causality is not "true causality, whatever that is," but he also maintains that his concept has a key role to play in the attempt to understand causality.

Although notions of cause and effect can be challenging to define in a general context, I believe it is possible to define satisfying notions of effect in experimental contexts. Generally, economists cannot conduct experiments in the way that laboratory scientists or clinical researchers can. Nevertheless, economists can often observe natural experiments, that is, identifiable discrete shifts in the economic environment, such as passage of a new law, deployment of a new invention or technology, implementation of a new economic or social policy, or a shift in industry behavior, such as a merger or the operation of a cartel. The effects of such natural experiments are often of keen interest to economists.

Economists have become increasingly sophisticated in their approach to measuring such effects A partial list of relevant work focusing on natural experiments is Angrist (2002a, 2002b), Auten and Carroll (1998), Berry and Waldfogel (2001), Bronars and Grogger (1994), Cook (2002), Deacon and Sonstelie (1985), Deltas and Kosmopoulou (2004), Frech (1976), Hotz, Mullin, and Sanders (1997), Kandel and Zilberfarb (1999), Metrick (1995), Meyer, Viscusi, and Durbin (1995), Millimet and List (2003), Rosenzweig and Wolpin (1980), Szymanski (2001), Treble (2003), and Zveglich and Rodger (2003). Useful methodological overviews directed toward the use (and mis-use) of natural experiments are given by Meyer (1995) and Rosenzweig and Wolpin (2000).

Despite the progress evident in this literature, it is still common for applied researchers to use the familiar and convenient "dummy variable" approach. To implement this approach in the regression context, the researcher simply includes a dummy variable equal to one for observations subject to the operation of the natural experiment and equal to zero otherwise. The dummy variable coefficient estimate is then interpreted as the effect of the natural experiment, holding all other factors constant (the "ceteris paribus effect").

As Higgins and Johnson (2003) note,

> An important practical use of dummy variables occurs in litigation when damages from some alleged misconduct is at issue. Here, it is common for the damage expert to find time periods or locations that differ according to the presence or absence of the complained of conduct. For example, in a price-fixing case, a cartel may be alleged to have reigned undetected in the 'before' period and, after detection, to have collapsed into some form of competition in the 'after' period. With sufficient observations on price and on supply and demand determinants in the two periods, a dummy variable approach might be used to estimate a 'monopoly overcharge'.

Not only "might" the dummy variable method be used, it is in fact extensively used in legal proceedings where damages depend on an estimate of what would have happened but for the alleged conduct. The work of Fisher (1980) and Rubinfeld (1985) in particular has had a lasting impact in this area. The dummy variable method has also found application in other areas with significant public policy import, such as merger event studies. There, the approach described above has been applied with the 'before' period corresponding to pre-merger observations and the 'after' period corresponding to post-merger observations, yielding, for example, estimates of the price effects of the merger. Examples are Vita and Sacher (2001), United States General Accounting Office (2004), and Taylor and Hosken (2004).

Given the common application of the dummy variable method in arenas with significant economic consequences, it is important to understand its properties as an estimator of the effect of a natural experiment. Accordingly, our first goal is to examine the conditions under which the dummy variable approach consistently estimates the effect of interest. We show that the required conditions are quite stringent. Higgins and Johnson (2003) provide conditions under which this approach gives an unbiased estimate of the desired effect. We extend their analysis not only to provide conditions under which the dummy variable method indeed delivers a consistent estimate, but also to examine in detail what happens when these ideal conditions fail.

Given the stringency of the conditions required for the dummy variable approach, our next goal is to provide alternative methods that will deliver consistent estimates of the desired effect under much less restrictive conditions. This approach is based on methods for estimating the "effect of treatment on the treated" in the treatment effects estimation literature (Rubin, 1974; Rosenbaum and Rubin, 1983; Hahn, 1998; Heckman, Ichimura, and Todd, 1998; Hirano, Imbens, and Ridder, 2003; see also Chen, Hong, and Tarozzi, 2004).

For an important class of natural experiments of interest here, the "treatment" is a structural change occurring at a given point in time, as in the cartel and merger event study examples. The treatment effect literature results do not immediately apply to such cases, however, as those results apply most readily to a cross-section of observational units (e.g., individuals) whose attributes (the "covariates") can be measured prior to the applied treatment and which therefore cannot possibly be affected by the treatment. In contrast, when a treatment occurs at some specific point in a time series, observations on the covariates may well occur subsequent to the treatment. Thus, particular care must be taken to insure that the covariates do not somehow embody effects of the treatment, as including such variables in the analysis can lead to serious distortions in the measured impact of the natural experiment. Accordingly, we contribute to the literature by providing in Section 2 an analytic framework in which one can (i) explicitly account for effects operating in time; and (ii) identify channels of indirect effect that may, if not properly handled, operate to introduce bias into the estimation of the effect of the natural experiment.

The framework of Section 2 does more than just provide a bridge between the existing treatment effects literature and treatments that operate in time. By focusing explicit attention on how the dependent variable of interest (e.g., the price of a product) responds to its determining variables (e.g., underlying cost and demand factors) and by requiring explicit recognition of which such variables are observable and which are not, our framework not only creates an explicit role for the operation of economic theory, but it also provides useful insight into how to construct the covariates. In the previous treatment effect literature, the desired effects are identified by simply assuming the existence of a set of covariates that satisfy Rubin's (1974) "unconfoundedness" assumption, a conditional independence condition related to a certain form of Granger non-causality. Our framework shows how such well-behaved covariates can be constructed as "predictive proxies," obtained as a combination of observable determining variables and observable proxies for unobservable determining variables. Thus, even for cross-section applications where existing results apply, the framework developed here provides useful insight into the selection of covariates.

Section 3 uses the foundation provided by Section 2 to analyze the properties of the dummy variable approach. In Section 4 we describe alternate procedures to estimate the effect of a natural experiment based on asymptotically efficient semiparametric methods previously proposed in the treatment effects literature by Hahn (1998) and by Hirano, Imbens, and Ridder (2003) (HIR).

Although the Hahn and HIR estimators are more generally valid in cross-section applications than the dummy variable method, they are computationally challenging (as they require non-parametric estimation) and their properties are unknown in the time series context of interest here. In Section 5 we propose a new computationally convenient "quasi-nonparametric" estimator that delivers an estimate of the effect of interest as the coefficient of a dummy variable for the treatment/natural experiment in a linear regression similar to the simple dummy variable approach of Section 3. We establish consistency and asymptotic normality of this estimator under conditions plausible for time series and we compare its efficiency to the Hahn and HIR estimators. Although this estimator is not necessarily asymptotically efficient, its computational convenience and known properties make it appealing for applications.

Given the important role played by the predictive proxies, we devote Sections 6 and 7 to a detailed examination of their construction. Section 7 also provides new tests of a key condition ("conditional exogeneity") that ensures unconfoundedness for a set of proposed predictive proxies. Section 8 contains concluding remarks.

2 The Data Generating Process and the Effects of Interest

A standard assumption in the treatment effect estimation literature is that the attributes of each observational unit are not affected by the treatment. For example, Rosenbaum and Rubin (1983, p. 42) state:

> Let x_i be a vector of observed pretreatment measurements or covariates for the ith unit; all of the measurements in x are made

prior to treatment assignment ...

By requiring the measurements to be made prior to treatment assignment, it is guaranteed that the treatment can have no impact on the covariates.

When the "treatment" of interest is a structural change occurring at a point in time, (e.g., the formation of a cartel or a merger event), the measurement of the covariates may well occur occur *subsequent* to the treatment. Particular care must be taken to ensure that the covariates do not somehow embody either direct or indirect effects of treatment. In this section, we set forth an analytic framework that permits us to explicitly account for effects operating in time and to identify channels of indirect effect that will require proper handling. This framework also provides the basis for defining the effects of interest here.

We begin with an economic theory specifying how the variable of interest, Y_t, is determined. In particular, we suppose this theory specifies

$$Y_t = c(\Lambda_t, \mathbf{Z}_t), \qquad t = 1, 2, \ldots,$$

where Λ_t is an indicator variable for the natural experiment, i.e. $\Lambda_t = 0$ in "regime 0" and $\Lambda_t = 1$ in "regime 1"; and $\mathbf{Z}_t$ is a vector of "determining variables" (whose elements may or may not be observable) specified by the theory. We view $\mathbf{Z}_t$ as stochastically generated by the operation of a relevant set of agents and markets, and we Λ_t as an intervention by some agent or collection of agents. The response function c is unknown.

Let $\mathcal{T}_0$ denote the regime 0 observation indexes, $\mathcal{T}_0 = \{t \in \mathbb{N} : \Lambda_t = 0\}$, and let $\mathcal{T}_1$ denote the regime 1 observation indexes, $\mathcal{T}_1 \equiv \{t \in \mathbb{N} : \Lambda_t = 1\}$. In time-series contexts, the observations for a given regime may or may not be contiguous. It is convenient to write

$$\begin{aligned} Y_t &= c_0(\mathbf{Z}_t) = c(0, \mathbf{Z}_t), \qquad t \in \mathcal{T}_0 \\ &= c_1(\mathbf{Z}_t) = c(1, \mathbf{Z}_t), \qquad t \in \mathcal{T}_1 \end{aligned}$$

to represent the operation of the natural experiment.

It is then natural to define the effect of the natural experiment as

$$\Delta(\mathbf{z}) = c_1(\mathbf{z}) - c_0(\mathbf{z}),$$

as this is the difference between the regimes of the natural experiment in outcomes observed for a fixed value $\mathbf{z}$ of the determining variables.

This measure is in fact central to our analysis, but particular care is required. When the determining variables may be generated subsequent to treatment, $\mathbf{Z}_t$ can be determined by Λ_t, in that economic theory entails

$$\mathbf{Z}_t = c^{(1)}(\Lambda_t, \mathbf{Z}_t^{(1)})$$

for some function $c^{(1)}$, say, and variables $\mathbf{Z}_t^{(1)}$, say, determining $\mathbf{Z}_t$.

For example, consider a time series example in which Y_t is the equilibrium price of some product of interest, $\Lambda_t = 1$ indicates the operation of an alleged cartel, and $\mathbf{Z}_t$ represents cost and demand shifters that determine the equilibrium price of the product. Then c_0 represents the non-cartel price determination relation (regime 0) and c_1 represents the cartel price determination relation (regime 1). Suppose that $\mathbf{Z}_t$ contains the price of a raw material, say $\mathbf{Z}_{1t}$, purchased by one member of the cartel from another. It is plausible that this price is determined differently in the cartel and non-cartel regimes, so $\mathbf{Z}_t = c^{(1)}(\Lambda_t, \mathbf{Z}_t^{(1)})$ with $c^{(1)}(0, \mathbf{Z}_t^{(1)}) \neq c^{(1)}(1, \mathbf{Z}_t^{(1)})$. The value of $\mathbf{Z}_{1t}$ will be impacted by the cartel, and $\Delta(\mathbf{z})$ thus will not measure the full effect of the cartel.

In the cartel example, interest attaches to what the price would have been in the absence of the cartel, which requires accounting not only for the direct effect of the cartel (through c) but also its indirect effect (through $\mathbf{Z}_t$). We have

$$\begin{aligned} Y_t &= c(\Lambda_t, c^{(1)}(\Lambda_t, \mathbf{Z}_t^{(1)})) \\ &= c^{[1]}(\Lambda_t, \mathbf{Z}_t^{(1)}), \end{aligned}$$

say. In such cases, and provided $\mathbf{Z}_t^{(1)}$ is not itself determined in some way by the cartel, the effect of interest, embodying both direct and indirect effects, is

$$\Delta^{[1]}(\mathbf{z}^{(1)}) = c_1^{[1]}(\mathbf{z}^{(1)}) - c_0^{[1]}(\mathbf{z}^{(1)}),$$

where $c_\lambda^{[1]}(\mathbf{z}^{(1)}) \equiv c^{[1]}(\lambda, \mathbf{z}^{(1)})$ for $\lambda = 0, 1$.

If, on the other hand, $\mathbf{Z}_t^{(1)}$ is non-trivially determined by Λ_t according to

$$\mathbf{Z}_t^{(1)} = c^{(2)}(\Lambda_t, \mathbf{Z}_t^{(2)}),$$

a further substitution gives

$$\begin{aligned} Y_t &= c^{[1]}(\Lambda_t, c^{(2)}(\Lambda_t, \mathbf{Z}_t^{(2)})) \\ &= c^{[2]}(\Lambda_t, \mathbf{Z}_t^{(2)}), \end{aligned}$$

The total effect accounting for direct effects and indirect effects is then

$$\Delta^{[2]}(\mathbf{z}^{(2)}) = c_1^{[2]}(\mathbf{z}^{(2)}) - c_0^{[2]}(\mathbf{z}^{(2)}).$$

By considering whatever deeper structures underlie the variables determining the target variable, we may eventually reach a set of markets and/or agents sufficiently removed from the operation of the natural experiment that it plays virtually no role in determining the variables at that level.

Let the first level of underlying determination with this property have index k. Then $\mathbf{Z}_t^{(k)}$ is neither directly nor indirectly determined by the natural experiment, and all the effects, whether direct or indirect, are captured by

$$\Delta^{[k]}(\mathbf{z}^{(k)}) = c_1^{[k]}(\mathbf{z}^{(k)}) - c_0^{[k]}(\mathbf{z}^{(k)}).$$

It is this total effect that will be the focus of our interest; so long as we understand that $\mathbf{Z}_t$ is neither directly nor indirectly affected by the natural experiment, then we can revert to our original notation and write

$$\Delta(\mathbf{z}) = c_1(\mathbf{z}) - c_0(\mathbf{z}).$$

To state our assumption on the data generating process in a way that accommodates this structure, we use the following definitions.

Definition 2.1 (Determining and Non-Determining Variables) *Let $\mathcal{V}, \mathcal{Z}_0$, and $\mathcal{Z}_1$ be random vectors such that $\mathcal{Z}_0 = f_0(\mathcal{V}, \mathcal{Z}_1)$ for some measurable function f_0. If in addition there exists a measurable function g_0 such that $f_0(\mathcal{V}, \mathcal{Z}_1) = g_0(\mathcal{Z}_1)$ almost surely (a.s.), then we say that $\mathcal{V}$ is* ***non-determining for*** *$\mathcal{Z}_0$* ***given*** *$\mathcal{Z}_1$, and we call $\mathcal{Z}_1$* ***the determining variables for*** *$\mathcal{Z}_0$.*

Definition 2.2 (Determining Chain; Isolatable and Isolating Variables) *Suppose there exists a random vector $\mathcal{V}$, random vectors $\mathcal{Z}_j$ and functions f_j, $j = 0, 1, \ldots$, such that*

$$\mathcal{Z}_j = f_j(\mathcal{V}, \mathcal{Z}_{j+1}), \qquad j = 0, 1, \ldots.$$

We call $(\mathcal{V}, \{\mathcal{Z}_j, f_j, j = 0, 1, \ldots\})$ a ***determining chain*** *for $\mathcal{Z}_0$. If there exists $k \in \mathbb{N}$ such that for all $j \geq k$ $\mathcal{V}$ is non-determining for $\mathcal{Z}_j$ given $\mathcal{Z}_{j+1}$, then we say that $\mathcal{V}$ is $(k-)$**isolatable** and that $\mathcal{Z}_k$* ***isolates*** *$\mathcal{V}$* ***for*** *$\mathcal{Z}_0$. We call $\mathcal{V}$ $(k-)$**isolatable variables**, and we call $\mathcal{Z}_k$* ***isolating variables for*** *$\mathcal{V}$. If $\mathcal{V}$ is 0-isolatable, we call $\mathcal{V}$* ***fundamentally non-determining for*** *$\mathcal{Z}_0$.*

To keep our discussion succinct, we mention only in passing that the notion of k-isolatable variables can be extended to permit degrees of isolation less than absolute. Heuristically, the idea is that at some remove, the isolated variables have an impact sufficiently small that they are "almost" isolated.

Our definitions permit us to consider relationships between target variables of interest $\mathcal{Z}_0$ and determining variables $\mathcal{V}$ and $\mathcal{Z}_k$ in which all the effects of $\mathcal{V}$, direct and indirect, are explicitly captured. Substituting gives

$$\begin{aligned}
\mathcal{Z}_0 &= f_0(\mathcal{V}, f_1(\mathcal{V}, \mathcal{Z}_1)) \\
&\equiv f_{[1]}(\mathcal{V}, \mathcal{Z}_1) \\
&= f_{[1]}(\mathcal{V}, f_2(\mathcal{V}, \mathcal{Z}_2)) \\
&\equiv f_{[2]}(\mathcal{V}, \mathcal{Z}_2) \\
&= \ldots \\
&= f_{[k]}(\mathcal{V}, \mathcal{Z}_k).
\end{aligned}$$

When $\mathcal{Z}_k$ isolates $\mathcal{V}$ for $\mathcal{Z}_0$, changes in $\mathcal{V}$ impact $\mathcal{Z}_0$ only through $f_{[k]}$ directly; there are no indirect effects, as $\mathcal{V}$ is fundamentally non-determining for $\mathcal{Z}_k$.

The function $f_{[k]}$ represents an analog of the reduced form of standard econometrics. To emphasize this, we call $f_{[k]}$ a "determining reduced form." One can analogously view f_0 (and $f_{[1]}, \ldots, f_{[k-1]}$) as a kind of structural equation, which we call a "determining structural equation." Just as standard systems of simultaneous equations can be analyzed using either the structural equations or the corresponding reduced form, so too can the effects of a natural experiment be analyzed through structural or reduced form analogs. For concreteness and simplicity, our focus here is exclusively on the determining reduced form. Measurement of the effects of natural experiments via determining structural equations permits an investigation of indirect effects, but for now, adequately treating the determining reduced form is enough to exhaust the space available.

By requiring that $\mathbf{Z}_t$ isolates Λ_t for Y_t, we ensure that $\mathbf{Z}_t$ constitutes a set of *sufficient concomitants* in the terminology of Dawid (2000), in that $\mathbf{Z}_t$ fully determines Y_t and $\mathbf{Z}_t$ is not determined by Λ_t. For convenience, however, we will refer to the elements of $\mathbf{Z}_t$ as "determining variables" (with the implicit understanding that Λ_t does not determine $\mathbf{Z}_t$) instead of as "concomitants," as the former nomenclature seems more intuitive.

The requirement that $\mathbf{Z}_t$ isolates Λ_t for Y_t now plays a key role in enforcing the analog of Rosenbaum and Rubin's (1983) requirement that the covariates be measured prior to treatment. The observable concomitants are useful covariates, so our requirment ensures that the natural experiment cannot impact them. Other covariates, whose construction we discuss at length below, will also be subject to the requirement that they are not determined by Λ_t.

For simplicity, we assume that the underlying determining chain for Y_t is the same for all t, and we say that $\mathbf{Z}_t$ ***stably*** *isolates* Λ_t *for* Y_t *according to*

$$Y_t = c(\Lambda_t, \mathbf{Z}_t), \qquad t = 1, 2, \ldots .$$

We will now leave implicit the determining chain and "isolation level," k.

Although our use of the subscript "t" to index observations is intended to suggest time series, our framework applies to both cross-section and time-series. Nevertheless, requiring that $\mathbf{Z}_t$ stably isolates Λ_t for Y_t imposes restrictions for time series that do not arise in cross sections. When an intervention lasts from period $\tau + 1$ to period n, we have

$$\Lambda_1 = \ldots = \Lambda_\tau = 0; \quad \Lambda_{\tau+1} = \ldots = \Lambda_n = 1.$$

It follows that the requirement that $\mathbf{Z}_t$ stably isolates Λ_t for Y_t rules out the presence of lagged values of Y_t in $\mathbf{Z}_t$. To see why, suppose for simplicity that

$$Y_t = c(\Lambda_t, Y_{t-1}),$$

so that $\mathbf{Z}_t = Y_{t-1}$. Then we also have

$$Y_{t-1} = c(\Lambda_{t-1}, Y_{t-2}) = c(\Lambda_t, Y_{t-2})$$

for all t except $t = \tau + 1$, so that Y_{t-1} cannot isolate Λ_t for Y_t.

Note that this restriction by no means rules out applications to time series generally. In the determining reduced form, the determining variables can contain current and lagged values of any time series not impacted by Λ_t.

We emphasize that this restriction is not a defect of the general approach taken here, but rather a consequence of the present focus on the determining reduced form. Treating dynamic determining relationships will require working with determining structural equations, but this is a task best deferred until the simpler task of understanding the reduced form has been accomplished.

We now have the concepts required to specify the data generating process.

Assumption A.1 (Data Generating Process) *The observed data are generated from a realization of the sequence of random variables* $(Y_t, \Lambda_t, \tilde{Z}_t, \ddot{Z}_t)$, $t = 1, 2 \ldots$, *where* $(\tilde{Z}_t, \ddot{Z}_t)$ *stably isolates* Λ_t *for* Y_t *according to*

$$Y_t = c(\Lambda_t, \tilde{Z}_t, \ddot{Z}_t), \qquad t = 1, 2, \ldots,$$

for some unknown measurable scalar-valued function c, *where* Λ_t *is* $\{0,1\}$-*valued. For* $i = 0, 1$, *define* $\mathcal{T}_i \equiv \{t \in \mathbb{N} : \Lambda_t = i\}$ *and assume that for all* $t \in \mathcal{T}_i$ $(\tilde{Z}_t, \ddot{Z}_t)$ *has joint distribution* F_i, $\ddot{Z}_t$ *has distribution* G_i, $\tilde{Z}_t$ *has joint distribution* H_i, *and the conditional distribution of* $\ddot{Z}_t$ *given* $\tilde{Z}_t = \tilde{z}$ *is* $\tilde{G}_i(\cdot \,|\, \tilde{z})$.

We refer to $\mathcal{T}_0$ and $\mathcal{T}_1$ simply as "regime 0" and "regime 1" respectively. We distinguish between $\tilde{Z}_t$ and $\ddot{Z}_t$ as follows. First, we treat $\tilde{Z}_t$ as observable and $\ddot{Z}_t$ as unobservable. Second, we view the elements of $\ddot{Z}_t$ unambiguously as determining variables for Y_t; however, we permit $\tilde{Z}_t$ to contain non-determining variables. As long as Λ_t has no impact on any non-determining variables (which is ensured because $(\tilde{Z}_t, \ddot{Z}_t)$ is isolating), then the presence of such non-determining variables has no adverse impact. Thus, we refer to $\ddot{Z}_t$ as the "unobservable determinants" of Y_t and to $\tilde{Z}_t$ as the "included observable" variables. We may for brevity refer to $(\tilde{Z}_t, \ddot{Z}_t)$ as "determinants" of Y_t, but with the understanding that some elements of $\tilde{Z}_t$ may be non-determining.

An interesting feature of the present framework is that although Λ_t has no direct or indirect impact on $(\tilde{Z}_t, \ddot{Z}_t)$, we nevertheless explicitly permit the joint distribution of $(\tilde{Z}_t, \ddot{Z}_t)$ to depend on Λ_t. We write

$$\begin{aligned} F_0(\tilde{z}, \ddot{z}) &= F(\tilde{z}, \ddot{z} \,|\, \Lambda_t = 0) \\ F_1(\tilde{z}, \ddot{z}) &= F(\tilde{z}, \ddot{z} \,|\, \Lambda_t = 1), \end{aligned}$$

where $F(\cdot \,|\, \Lambda_t)$ is the conditional distribution of $(\tilde{Z}_t, \ddot{Z}_t)$ given Λ_t. In this context, $F_0 = F_1$ is equivalent to independence of Λ_t and $(\tilde{Z}_t, \ddot{Z}_t)$. But there is nothing in our requirements that can guarantee $F_0 = F_1$.

Indeed, economists should expect $F_0 \neq F_1$, as whatever processes act to generate $(\tilde{Z}_t, \ddot{Z}_t)$ can easily do so in such a way that their joint distribution differs between regimes. For example, suppose that the price of natural gas is not impacted by a cartel either directly or indirectly and that it is a determining variable. Forces in the economy can easily yield distributions of natural gas prices that differ between cartel and non-cartel regimes. For example, more extreme weather during the cartel regime could cause the average price of natural gas to be higher in the cartel regime.

Indeed, disentangling the effects of the natural experiment from the effects of other factors for which $F_0 \neq F_1$ is one of the most critical challenges in measuring the effect of a natural experiment. A major focus of our attention will thus be on handling the implications of $F_0 \neq F_1$, with particular attention directed to the separate implications of $H_0 \neq H_1$ and $\tilde{G}_0 \neq \tilde{G}_1$.

Proceeding now to consider the effect of the natural experiment, we see that the relationship between Y_t and its determinants changes with the regime shift:

$$\begin{aligned} Y_t &= c_0(\tilde{Z}_t, \ddot{Z}_t), \qquad t \in \mathcal{T}_0 \\ Y_t &= c_1(\tilde{Z}_t, \ddot{Z}_t), \qquad t \in \mathcal{T}_1. \end{aligned}$$

Thus, the *effect of the natural experiment given* $(\tilde{z}, \ddot{z})$ is

$$\Delta(\tilde{z}, \ddot{z}) \equiv c_1(\tilde{z}, \ddot{z}) - c_0(\tilde{z}, \ddot{z}).$$

Note that by virtue of the fact that $(\tilde{Z}_t, \ddot{Z}_t)$ isolates Λ_t, this is the total effect of the natural experiment, capturing all effects, whether direct or indirect.

We pay particular attention to certain average effects. For example, let $\tilde{G}(\cdot|\tilde{z})$ be a conditional distribution for $\ddot{Z}_t$ given $\tilde{Z}_t = \tilde{z}$, and define

$$\tilde{\Delta}(\tilde{z}, \tilde{G}) \equiv \int \Delta(\tilde{z}, \ddot{z}) \, d\tilde{G}(\ddot{z}|\tilde{z}).$$

This is the *average effect of the natural experiment given* $(\tilde{z}, \tilde{G})$, with the average taken over the unobservable determinants, given the observable determinants.

Next, letting $\tilde{G}(\cdot|\tilde{z})$ be some conditional distribution for $\ddot{Z}_t$ given $\tilde{Z}_t = \tilde{z}$ and H be some joint distribution for $\tilde{Z}_t$, define

$$\Delta^*(H, \tilde{G}) \equiv \int \tilde{\Delta}(\tilde{z}; \tilde{G}) \, dH(\tilde{z}).$$

$\Delta^*(H, \tilde{G})$ is the *average effect of the natural experiment given* $(H, \tilde{G})$. Of particular interest is the case in which $H = H_1$ and $\tilde{G} = \tilde{G}_1$; we write

$$\Delta_1^* \equiv \Delta^*(H_1, \tilde{G}_1).$$

In the treatment effects literature, this is the "average treatment effect on the treated." As this will be a main focus of interest for reasons to be elaborated next, we will simply refer to this as the "effect of interest."

In the cartel example, Y_t is the equilibrium price of a product and $\tilde{Z}_t$ and $\ddot{Z}_t$ are cost and demand shifters not affected by the cartel and that determine equilibrium price. Recall that c_0 represents the non-cartel price-determination relationship, and c_1 represents the cartel price-determination relationship. Then $\Delta(\tilde{z}, \ddot{z})$ represents the difference in prices between the collusive and non-collusive regimes, for any given configuration of all demand and cost shifters ("market conditions"). This effect is not observable, as $\ddot{Z}_t$ is unobservable.

In contrast, $\tilde{\Delta}(\tilde{z}, \tilde{G}_1)$ is the average price difference (collusive effect) occurring given observable market conditions $\tilde{Z}_t = \tilde{z}$, averaged over the conditional distribution $\tilde{G}_1$ of the unobservables of the collusive regime given $\tilde{Z}_t = \tilde{z}$. Similarly, Δ_1^* is the (unconditional) average price difference (collusive effect) under the average market conditions of the collusive regime, as the average is over all price determinants, according to the joint distribution prevailing during the collusive regime.

We have

$$\begin{aligned}\tilde{\Delta}(\tilde{z}, \tilde{G}_1) &= \int \Delta(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z}) \\ &= \int [c_1(\tilde{z}, \ddot{z}) - c_0(\tilde{z}, \ddot{z})]\, d\tilde{G}_1(\ddot{z}|\tilde{z}) \\ &= \tilde{\mu}_1(\tilde{z}) - \tilde{\mu}_{01}(\tilde{z}),\end{aligned}$$

with

$$\begin{aligned}\tilde{\mu}_1(\tilde{z}) &\equiv \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z}) \\ \tilde{\mu}_{01}(\tilde{z}) &\equiv \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z}).\end{aligned}$$

The quantity $\tilde{\mu}_{01}(\tilde{z})$ is especially important in antitrust analysis, as it is the expected "but-for" price, that is, the price that would have been expected to have occurred given observable market conditions $\tilde{z}$, had the non-collusive pricing relation operated under the market conditions prevailing in the collusive regime, taking into account all effects of the cartel, whether direct or indirect.

Similarly, we have

$$\begin{aligned}\Delta_1^* &= \int \Delta(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\ &= \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) - \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\ &= \mu_1 - \mu_{01},\end{aligned}$$

with

$$\begin{aligned}\mu_1 &\equiv \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) = \int \tilde{\mu}_1(\tilde{z})\ dH_1(\tilde{z}) \\ \mu_{01} &\equiv \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) = \int \tilde{\mu}_{01}(\tilde{z})\ dH_1(\tilde{z}).\end{aligned}$$

Thus, the average effect of the collusion, Δ_1^*, is the difference between μ_1, the average collusive price in the cartel regime, and μ_{01}, the average but-for price in the cartel regime. Whereas $\tilde{\Delta}(\tilde{z}, \tilde{G}_1)$ gives a measure of impact conditional on $\tilde{Z}_t = \tilde{z}$, Δ_1^* provides a measure of overall (unconditional) average impact.

3 The Dummy Variable Approach

As noted in the introduction, a common approach to attempting to measure the effect of a natural experiment is to estimate the "dummy variable model,"

$$Y_t = \Lambda_t \alpha + Z_t' \beta + v_t, \qquad t = 1, 2, \ldots,$$

where Λ_t is the dummy (indicator) for the natural experiment: $\Lambda_t = 0$ for $t \in \mathcal{T}_0$, $\Lambda_t = 1$ for $t \in \mathcal{T}_1$; $Z_t = (1, \tilde{Z}_t)'$ is a column vector of included regressors, and v_t is a "residual," intended to accommodate the effect on Y_t of $\ddot{Z}_t$, the unobservable determinants of Y_t.

For simplicity, we assume the model is estimated using ordinary least squares (OLS). The OLS parameter estimator for a sample of size n is

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \Lambda'\Lambda & \Lambda'Z \\ Z'\Lambda & Z'Z \end{pmatrix}^{-1} \begin{pmatrix} \Lambda'Y \\ Z'Y \end{pmatrix},$$

where Λ is the $n \times 1$ vector with elements Λ_t, Z is the $n \times k$ matrix with rows Z_t', and Y is the $n \times 1$ vector with elements Y_t. Standard manipulations permit us to write $\hat{\alpha}$ and $\hat{\beta}$ in terms of sample moments as follows:

$$\begin{aligned} \hat{\alpha} &= \hat{\mu}_1 - \hat{m}_1'\hat{\beta} \\ \hat{\beta} &= [(1-\hat{p}_1)\hat{M}_0 + \hat{p}_1\hat{M}_1 - \hat{p}_1\hat{m}_1\hat{m}_1']^{-1} \\ &\quad \times[(1-\hat{p}_1)\hat{L}_0 + \hat{p}_1\hat{L}_1 - \hat{p}_1\hat{m}_1\hat{\mu}_1], \end{aligned}$$

where the proportion of observations in regime 1 is $\hat{p}_1 \equiv T_1/n$, with T_1 the number of regime 1 observations, $T_1 \equiv \sum_{t=1}^n \Lambda_t$, $T_0 \equiv n - T_1$ is the number of regime 0 observations, and the sample moments are

$$\begin{aligned} \hat{\mu}_1 &\equiv T_1^{-1}\sum_{t=1}^n \Lambda_t Y_t, \qquad \hat{m}_1 \equiv T_1^{-1}\sum_{t=`}^n \Lambda_t Z_t, \\ \hat{M}_0 &\equiv T_0^{-1}\sum_{t=1}^n (1-\Lambda_t) Z_t Z_t', \qquad \hat{M}_1 \equiv T_1^{-1}\sum_{t=1}^n \Lambda_t Z_t Z_t', \\ \hat{L}_0 &\equiv T_0^{-1}\sum_{t=1}^n (1-\Lambda_t) Z_t Y_t, \qquad \hat{L}_1 \equiv T_1^{-1}\sum_{t=1}^n \Lambda_t Z_t Y_t. \end{aligned}$$

According to the standard textbook interpretation of OLS (e.g. Greene, 1993, pp. 231-232), $\hat{\alpha}$ estimates the ceteris paribus effect of the natural experiment (the effect of interest) and $\hat{\beta}$ estimates the ceteris paribus effects of all the other included determinants. We now consider whether or not the textbook interpretation is generally valid or plausible. We do this by giving general conditions ensuring that $\hat{\alpha}$ and $\hat{\beta}$ converge to well-defined probability limits and then interpreting these probability limits

It is straightforward to give conditions ensuring that $\hat{\alpha}$ and $\hat{\beta}$ converge to well-defined probability limits. The following assumptions suffice.

Assumption A.2 (Finiteness of Moments) *Let* $z \equiv (1, \tilde{z})'$.

(a) $0 < p_1 < 1$;

(b) (i) $M_0 \equiv \int zz' dH_0(\tilde{z}) < \infty, \quad \det M_0 > 0$;

(ii) $M_1 \equiv \int zz' dH_1(\tilde{z}) < \infty, \quad \det M_1 > 0$;

(c) (i) $L_0 \equiv \int z c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z}) < \infty$;

(ii) $L_1 \equiv \int z c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) < \infty$.

This assumption is mild and thus plausible. Assumption A.2(a) ensures that when $\hat{p}_1 \xrightarrow{p} p_1$, then both regimes constitute a non-vanishing proportion of the data. Assumption A.2(b) ensures that the regressors have finite and non-singular second moments in each regime, and A.2(c) ensures that the regressor and dependent variable cross-moments are finite in each regime.

Assumption A.3 (Laws of Large Numbers)

(a) $\hat{p}_1 \xrightarrow{p} p_1$;

(b) $\hat{M}_0 \xrightarrow{p} M_0$ and $\hat{M}_1 \xrightarrow{p} M_1$;

(c) $\hat{L}_0 \xrightarrow{p} L_0$ and $\hat{L}_1 \xrightarrow{p} L_1$.

These law of large number requirements are also mild and hold under a wide variety of differing primitive conditions on $\{\Lambda_t, \tilde{Z}_t, \ddot{Z}_t\}$ suitable either for cross-section or time-series data (see, e.g., White, 2001, ch. 3). Assumptions A.2 and A.3 can be further relaxed, but we maintain them for simplicity.

Our first result follows immediately from the continuity of $\hat{\alpha}$ and $\hat{\beta}$ as functions of the sample moments and their convergence in probability.

Proposition 3.1 *Given A.1 -A.3, $\hat{\alpha} \xrightarrow{p} \alpha^*$ and $\hat{\beta} \xrightarrow{p} \beta^*$, where*

$$\begin{aligned}
\alpha^* &\equiv \mu_1 - m_1'\beta^*, \\
\beta^* &\equiv [(1-p_1)M_0 + p_1 M_1 - p_1 m_1 m_1']^{-1} \\
&\quad \times [(1-p_1)L_0 + p_1 L_1 - p_1 m_1 \mu_1],
\end{aligned}$$

with $m_1 \equiv \int z\, dH_1(\tilde{z})$.

Proofs are given in the Mathematical Appendix.

Our primary focus is on α^*, as this has the standard interpretation as the effect of interest, i.e., the ceteris paribus effect of the natural experiment. Recall that the average effect of the treatment on the treated is

$$\Delta_1^* = \mu_1 - \mu_{01}.$$

Comparing α^* and Δ_1^*, we see that a necessary and sufficient condition that $\alpha^* = \Delta_1^*$, so that α^* gives the effect of interest, is that

$$\mu_{01} = m_1'\beta^*.$$

Accordingly, we now investigate the relation between α^* and Δ_1^* (equivalently μ_{01} and $m_1'\beta^*$). The following result permits a comprehensive analysis.

Proposition 3.2 *Suppose Assumptions A.1 and A.2 hold and that in addition* $L_{01} \equiv \int z c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) < \infty$. *Then*

$$\begin{aligned}
\alpha^* - \Delta_1^* &= m_1'(\beta_{01}^* - \beta_0^*) + p_1(m_0 - m_1)' S' S(\beta_1^* - \beta_0^*) \\
&\quad + p_1(1-p_1)(m_0 - m_1)' S' \tilde{M}^{-1}(\tilde{M}_1 - \tilde{M}_0) S(\beta_1^* - \beta_0^*) \\
&= \int c_0(\tilde{z}, \ddot{z})(d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z}))\, dH_1(\tilde{z}) \\
&\quad + \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, (dH_1(\tilde{z}) - dH_0(\tilde{z})) \\
&\quad + p_1(m_0 - m_1)'\beta_1^* + (1-p_1)(m_0 - m_1)'\beta_0^* \\
&\quad + p_1(1-p_1)(m_0 - m_1)' S' \tilde{M}^{-1}(\tilde{M}_1 - \tilde{M}_0) S(\beta_1^* - \beta_0^*),
\end{aligned}$$

where $m_0 \equiv \int z\ dH_0(\tilde{z})$, $\beta_0^* \equiv M_0^{-1} L_0$, $\beta_1^* \equiv M_1^{-1} L_1$, $\beta_{01}^* \equiv M_1^{-1} L_{01}$. S is *the selection matrix that selects the non-constant elements of* z ($\tilde{z}' = Sz$), *and*

$$\begin{aligned}
\tilde{M} &\equiv (1-p_1)\tilde{M}_0 + p_1 \tilde{M}_1, \\
\tilde{M}_0 &\equiv S(M_0 - m_0 m_0')S', \\
\tilde{M}_1 &\equiv S(M_1 - m_1 m_1')S'.
\end{aligned}$$

This gives us two contrasting conditions that make it straightforward to identify circumstances in which the "apparent effect" α^* coincides with the effect of interest, Δ_1^*. Further, these expressions enable us to examine the various roles played by different aspects of the DGP in creating a "causal discrepancy" between the apparent effect and the effect of interest.

The second expression provides the simplest insight. From this expression we see that when $H_0 = H_1$ and $\tilde{G}_0 = \tilde{G}_1$, then each of the five discrepancy terms vanishes. Thus, the sufficient conditions for $\alpha^* = \Delta_1^*$ evident from the second expression is that the joint distributions F_0 and F_1 of the included observables ($\tilde{Z}_t$) and the unobservable determinants ($\ddot{Z}_t$) are identical across regimes. We state this result formally as follows:

Corollary 3.3 *Suppose Assumptions A.1 and A.2 hold and that* $L_{01} < \infty$. *If in addition* $H_0 = H_1$ *and* $\tilde{G}_0 = \tilde{G}_1$, *a.s.* $- H_1 (= H_0)$, *then* $\alpha^* = \Delta_1^*$.

Note that $\alpha^* = \Delta_1^*$ without the dummy variable model having to be correctly specified, i.e., we do not require that for some δ^*

$$c_1(\tilde{z}, \ddot{z}) = \delta^* + c_0(\tilde{z}, \ddot{z}).$$

Note also that the conditions $H_0 = H_1$, $\tilde{G}_0 = \tilde{G}_1$ are precisely what a laboratory scientist (an "experimenter") or clinical researcher can achieve by assigning treatment at random ("randomization").

In certain circumstances it may be possible to achieve $H_0 = H_1$ and $\tilde{G}_0 = \tilde{G}_1$ even without experimental control. This can occur in studies of twins (e.g. Bronars and Grogger (1994) and Rosenzweig and Wolpin (1980)), although here one may well find $H_0 \neq H_1$ due to varying life circumstances. As it may be possible to accomodate $H_0 \neq H_1$ by suitable regression methods, this need not be problematic.

Generally, however, there is nothing to ensure that nature, in conducting natural experiments, does so in such a way as to assign treatment to cases at random. Economists who cannot control the phenomena they observe (and who lack samples of twins) must confront the fact that in general either or, more likely, both $H_0 = H_1$ and $\tilde{G}_0 = \tilde{G}_1$ may fail. Because of this lack of experimental control, we henceforth refer to such economists or other similarly experimentally challenged researchers as "observers."

Indeed, not only should economists not expect nature to perform experiments in the way that they might wish, but economists should expect that other economists may conduct analyses, especially when the effect of the natural experiment may be controversial, so as to guarantee that $m_0 \neq m_1$. This is well illustrated by the case in which antitrust damages are to be calculated. If increases in price are the result of demonstrable changes in market conditions (for example, increases in the average levels of cost and demand shifters), then increases in price can be explained as the innocent effect of these changes, rather than the illegal operation of a cartel. In such cases, one should expect analyses performed on behalf of defendants in antitrust proceedings to point to precisely such supposedly innocent effects, necessitating $m_0 \neq m_1$. Proposition 3.2 then applies, and, as should be evident from this result, the researcher's discretion over which observable variables to include (and which to exclude) when using the dummy variable approach creates an expansive opportunity for the manipulation of results to favor one side or the other.

Given that economists and observers generally cannot rely on nature (or colleagues in controversy) to conveniently arrange matters so that $H_0 = H_1$ and $\tilde{G}_0 = \tilde{G}_1$, we ask whether other conditions may yet guarantee $\alpha^* = \Delta_1^*$. Such conditions are readily extracted from the first expression for $\alpha^* - \Delta_1^*$ in Proposition 3.2, from which it suffices that $\beta_{01}^* = \beta_0^*$ and $S(\beta_1^* - \beta_0^*) = 0$. The first condition causes the first term $m_1'(\beta_{01}^* - \beta_0^*)$ to vanish, whereas the second causes the remaining two terms to vanish.

The following result shows what is required to deliver these conditions, when the dummy variable model is correctly specified to a high degree.

Corollary 3.4 *Suppose Assumptions A.1 and A.2 hold and that* $L_{01} < \infty$.

(i) *Suppose further that*

$$c_0(\tilde{z}, \ddot{z}) = z'b^* + u_0(\ddot{z})$$

for some unknown finite vector b^* and measurable function u_0, such that

$$\int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z}) = 0,$$

and that for some unknown finite scalar δ^* we have

$$c_1(\tilde{z}, \ddot{z}) = \delta^* + c_0(\tilde{z}, \ddot{z}).$$

Then

$$\begin{aligned}
\beta_0^* &= b^*, \\
\beta_{01}^* &= b^* + M_1^{-1} \int z u_0(\ddot{z}) \left[d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z}) \right] dH_1(\tilde{z}), \\
\beta_1^* &= \beta_{01}^* + S_1' \delta^*,
\end{aligned}$$

where $S_1 = (1, 0, \ldots, 0)$.

(ii) *If in addition $\tilde{G}_0 = \tilde{G}_1$ a.s. $- H_1$, then $\beta_0^* = \beta_{01}^*$ and $S(\beta_1^* - \beta_0^*) = 0$, so that*

$$\alpha^* = \delta^* = \Delta_1^*.$$

This says that the dummy variable approach consistently estimates the effect of interest, provided that the model is fully correctly specified and that $\tilde{G}_0 = \tilde{G}_1$.

By assuming that $c_0(\tilde{z}, \ddot{z}) = z'b^* + u_0(\ddot{z})$ and $c_1(\tilde{z}, \ddot{z}) = \delta^* + c_0(\tilde{z}, \ddot{z})$, we have the DGP

$$Y_t = \Lambda_t \delta^* + Z_t b^* + \upsilon_t, \qquad t = 1, 2, \ldots,$$

This is an extremely strong assumption, as it not only imposes the restriction that the effect of the natural experiment is an additive shift δ^*, but it also requires: (i) that the effects of the remaining underlying variables are separable between observables and unobservables; and (ii) that the effects of the included observable variables must be separable and proportional. Both of these are strong and not particularly plausible assumptions.

A standard assumption economists employ to ensure $\tilde{G}_0 = \tilde{G}_1$ is to assume that $(\Lambda_t, \tilde{Z}_t)$ is independent of $\ddot{Z}_t$. In this case, we have

$$\tilde{G}_0(\ddot{z}|\tilde{z}) = \tilde{G}_1(\ddot{z}|\tilde{z}) = G(\ddot{z}).$$

Significantly, the condition $\tilde{G}_0 = \tilde{G}_1$ does not require independence of $(\Lambda_t, \tilde{Z}_t)$ and $\ddot{Z}_t$. For $\lambda = 0, 1$, write

$$\tilde{G}(\ddot{z}|\lambda, \tilde{z}) = \tilde{G}_\lambda(\ddot{z}|\tilde{z}).$$

The requirement that $\tilde{G}_0 = \tilde{G}_1$ can then be expressed as, say,

$$\tilde{G}(\ddot{z}|\lambda, \tilde{z}) = \tilde{G}(\ddot{z}|\tilde{z}), \qquad \lambda = 0, 1,$$

indulging in a mild abuse of notation. This expresses the property that $\ddot{Z}_t$ is independent of Λ_t given $\tilde{Z}_t$, written

$$\ddot{Z}_t \perp \Lambda_t \mid \tilde{Z}_t.$$

This *conditional independence* property plays a key role in statistics (Dawid, 1979), and is closely related to the concept of Granger (1969) non-causality. (See Florens and Mouchart (1982) and Florens and Fougere (1996) for detailed discussion of this relation.) Together with correct specification, this conditional independence condition ensures $\alpha^* = \Delta_1^* = \delta^* = \Delta(\tilde{z}, \ddot{z})$, so that the dummy variable method consistently estimates the desired effects.

Clinical investigators can ensure conditional independence by assigning treatment independently of the unobservables, as we then have

$$d\tilde{G}_0(\ddot{z}|\tilde{z}) = d\tilde{G}_1(\ddot{z}|\tilde{z}) = d\tilde{G}(\ddot{z}|\tilde{z}).$$

The use of twins may also achieve this goal. Note that even if the investigator or twins researcher has $H_0 \neq H_1$, the correct specification of the model neutralizes any adverse impact for $\alpha^* = \Delta_1^*$.

The result of 3.4(ii) does not, however, offer much comfort to economists or observers generally, as nature cannot plausibly be trusted to ensure the desired conditional independence, $\tilde{G}_0 = \tilde{G}_1$. The conditions of Corollary 3.4 also require correct specification to a high degree. Our next result shows that correct specification is a necessary condition when $H_0 \neq H_1$, as α^* is no longer consistent for Δ_1^* when the dummy variable model is misspecified in even a modest way.

Corollary 3.5 *Suppose Assumptions A.1 and A.2 hold and that* $L_{01} < \infty$.

(i) *Suppose further that*

$$\begin{aligned} c_0(\tilde{z}, \ddot{z}) &= z' b_0^* + u_0(\ddot{z}) \\ c_1(\tilde{z}, \ddot{z}) &= z' b_1^* + u_1(\ddot{z}) \end{aligned}$$

for unknown finite vectors b_0^* *and* b_1^* *and measurable functions* u_0 *and* u_1 *such that*

$$\begin{aligned} \int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z}) &= 0 \\ \int u_1(\ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z}) &= 0. \end{aligned}$$

Then

$$\begin{aligned} \beta_0^* &= b_0^* \\ \beta_{01}^* &= b_0^* + M_1^{-1} \int z u_0(\ddot{z}) [d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z})]\, dH_1(\tilde{z}) \\ \beta_1^* &= b_1^*. \end{aligned}$$

(ii) *If in addition* $\tilde{G}_0 = \tilde{G}_1\, a.s. - H_1$, *then* $\beta_0^* = \beta_{01}^*$ *and*

$$\begin{aligned}\alpha^* - \Delta_1^* &= p_1(m_0 - m_1)'S'S(b_1^* - b_0^*) \\ &\quad + p_1(1-p_1)(m_0 - m_1)S'\tilde{M}^{-1}(\tilde{M}_1 - \tilde{M}_0)S(b_1^* - b_0^*).\end{aligned}$$

We characterize the misspecification as "mild" in that the response functions c_0 and c_1 are still separable between the observables and unobservables and linear (actually affine) in the included observable variables. Now, however, the ceteris paribus effect of the natural experiment is not simply an additive shift δ^*, but is instead $\Delta(\tilde{z}, \ddot{z}) = z'(b_1^* - b_0^*) + u_1(\ddot{z}) - u_0(\ddot{z})$. The natural experiment has the effect of changing the way that the dependent variable responds to the underlying variables.

For cases like the natural experiment of a cartel, this situation is highly relevant, as economic theory dictates that the way that equilibrium prices respond to cost and demand shifters generally depends on the degree of competition in the industry. For example, perfect competitors pass along 100% of cost increases, whereas a monopoly (a perfect cartel) passes along less than 100% of cost increases.

In this case, the effect of interest is

$$\Delta_1^* = m_1'(b_1^* - b_0^*) + \int u_0(\ddot{z})\, dG_1(\ddot{z}).$$

But even if the observer were to somehow achieve $\tilde{G}_0 = \tilde{G}_1$, which then yields $\Delta_1^* = m_1'(b_1^* - b_0^*)$, we see from 3.5(ii) that the combination of $H_0 \neq H_1$ and $S(b_1^* - b_0^*) \neq 0$ creates a wide field of possibilities for the causal discrepancy $\alpha^* - \Delta_1^*$, readily subject to manipulation by interested parties.

Note that Corollary 3.5(ii) contains as a special case the analog of the first Proposition of Higgins and Johnson (2003, p. 258), in which they establish the unbiasedness of the dummy coefficient as an estimate of the average effect when the linear model is correctly specified (although with differing slopes), but the included regressors have identical means ($m_0 = m_1$). If indeed $m_0 = m_1$ with correct specification and conditional independence of $\ddot{Z}_t$ and Λ_t given $\tilde{Z}_t$ ($\tilde{G}_0 = \tilde{G}_1$) then the dummy coefficient is consistent for the desired effect as a consequence of Corollary 3.5 (ii).

The message from this analysis is that economists cannot generally rely on the simple dummy variable approach to measure the effect of interest Δ_1^* for a natural experiment, much less the ceteris paribus effect $\Delta(\tilde{z}, \ddot{z})$. In the hands of economists, the dummy variable approach is fatally flawed, as it can be used to manipulate the apparent effect (α^*) of a natural experiment to support a wide range of different positions. The reason for this is the economist's lack of experimental control. Even if the dummy variable model were correctly specified as a description of the natural experiment at hand, the economist is not able to ensure $\tilde{G}_0 = \tilde{G}_1$ in the same way that an investigator or twins researcher can. The economist thus cannot plausibly appeal to Corollary 3.4 to claim that the causal discrepancy vanishes. Moreover, when the dummy variable model is misspecified, as is generally the case, the economist again cannot exercise experimental control in the same way that an experimenter or investigator can (randomizing treatment so that $\tilde{G}_0 = \tilde{G}_1$ and $H_0 = H_1$) to overcome the otherwise adverse impact of the misspecification. Except in highly exceptional circumstances, the dummy variable approach yields an apparent effect α^* of a natural experiment that can be considerably at variance with the effect of interest, Δ_1^*.

4 The Treatment Effects Approach

The results of Section 3 show that the dummy variable approach to measuring the effect of a natural experiment is valid only under very strong assumptions. Although experimenters and, in certain circumstances, clinical investigators can ensure that these assumptions hold by virtue of their ability to randomize treatment, economists and observers generally cannot rely on nature to conduct experiments with the required regard for the researcher's objectives. Thus, the econometric folklore justifying the dummy variable approach is misleading, and another approach must be found.

In analyzing the dummy variable approach we essentially considered whether given statistical methods (dummy variable regression) were capable of achieving a specified goal (consistent estimation of Δ_1^*). Here we take a direct approach and ask what statistical methods can be brought to bear to achieve our goal.

Thus, consider estimating the average effect of the natural experiment,

$$\Delta_1^* = \mu_1 - \mu_{01},$$

where we recall that

$$\begin{aligned}
\mu_1 &\equiv \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
\mu_{01} &\equiv \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}).
\end{aligned}$$

It is easy to construct a consistent estimator for μ_1. For example, one can use the regime 1 average,

$$\hat{\mu}_1 = T_1^{-1} \sum_{t=1}^{n} \Lambda_t Y_t.$$

The challenge is to estimate μ_{01} consistently. We saw in Section 3 that only in very special circumstances is the estimator $\hat{m}_1' \hat{\beta}$ given by the dummy variable approach useful for this purpose. Accordingly we seek a method that, unlike the dummy variable approach, will not be adversely affected by model misspecification and the failure of independence or conditional independence.

A way forward is provided by the properties of conditional expectation. Suppose that in addition to $\tilde{Z}_t$, we can also observe auxiliary variables $\tilde{W}_t$, which we view as proxies for the unobservable determinants $\ddot{Z}_t$. Note that the $\tilde{W}_t$'s are not themselves determinants of Y_t, as they are observable, and all observable determinants have already been included in $\tilde{Z}_t$.

Let the joint distributions of $(\tilde{Z}_t, \tilde{W}_t)$ be denoted $\breve{H}_0$ and $\breve{H}_1$ for regimes 0 and 1 respectively, and let the conditional distributions of $\ddot{Z}_t$ given $(\tilde{Z}_t, \tilde{W}_t) = (\tilde{z}, \tilde{w})$ be given by $\breve{G}_0(\cdot|\tilde{z}, \tilde{w})$ and $\breve{G}_1(\cdot|\tilde{z}, \tilde{w})$ for regimes 0 and 1 respectively. We view $\breve{G}_0$ and $\breve{G}_1$ as the predictive distributions of $\ddot{Z}_t$ given $(\tilde{Z}_t, \tilde{W}_t)$, and we accordingly refer to $(\tilde{Z}_t, \tilde{W}_t)$ jointly as "predictive proxies" for $\ddot{Z}_t$.

With this notation and using the properties of conditional expectation, we have

$$\begin{aligned} \mu_{01} &= \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\ &= \int c_0(\tilde{z}, \ddot{z})\, d\breve{G}_1(\ddot{z}|\tilde{z}, \tilde{w})\, d\breve{H}_1(\tilde{z}, \tilde{w}). \end{aligned}$$

The key to the treatment effect approach is to exploit economic theory to choose $\tilde{W}_t$ in such a way that even though we may not have $\tilde{G}_0 = \tilde{G}_1$, we do have

$$\breve{G}_0 = \breve{G}_1 \qquad (a.s. - \breve{H}_1).$$

Just as $\tilde{G}_0 = \tilde{G}_1$ represents the conditional independence property

$$\ddot{Z}_t \perp \Lambda_t \mid \tilde{Z}_t,$$

the condition $\breve{G}_0 = \breve{G}_1$ represents the conditional independence property

$$\ddot{Z}_t \perp \Lambda_t \mid (\tilde{Z}_t, \tilde{W}_t),$$

which for convenience we refer to as "conditional independence given predictive proxies," or simply CIPP. When CIPP holds, we say that the predictive proxies ensure the "conditional exogeneity" of the treatment, or even more simply, that conditional exogeneity holds. The modifier "conditional" serves a reminder that the referenced property involves conditioning. The designation "exogenous" indicates that certain otherwise problematic dependencies between the treatment and the unobservables are absent.

The CIPP condition is closely related to Rubin's (1974) "unconfoundedness" assumption, also known as "selection on observables." To draw the necessary connections in a notationally compatible way, let Λ be the treatment indicator ($\Lambda = 1$ with treatment, $\Lambda = 0$ otherwise), let Y^0 denote the outcome without treatment, let Y^1 denote the outcome with treatment, and let X denote covariates not impacted by the treatment. Assume that given any possible value x of X, both treatment and non-treatment occur with positive probability:

$$0 < P[\Lambda = 1 | X = x] < 1.$$

Our interest here attaches to the average treatment effect on the treated,

$$\Delta_1^* \equiv E(Y^1 - Y^0 | \Lambda = 1).$$

Applying the law of iterated expectations gives

$$\Delta_1^* = E(E(Y^1|\Lambda = 1, X)|\Lambda = 1) - E(E(Y^0|\Lambda = 1, X)|\Lambda = 1).$$

Rubin's unconfoundedness assumption is that

$$(Y^0, Y^1) \perp \Lambda \mid X.$$

With this condition we have

$$E(Y^0|\Lambda = 1, X) = E(Y^0|\Lambda = 0, X) = E(Y^0|X),$$

so we can substitute in the second term of the above expression for Δ_1^* to obtain

$$\Delta_1^* = E(E(Y^1|\Lambda = 1, X)|\Lambda = 1) - E(E(Y^0|\Lambda = 0, X)|\Lambda = 1).$$

Only the second term presents a challenge, as the first term is just $E(Y^1|\Lambda = 1)$. But $E(Y^0|X) = E(Y^0|\Lambda = 0, X)$ can now be identified from observations on the untreated cases, ensuring that the effect of interest Δ_1^* can be identified. (see, e.g. Heckman, Ichimura, and Todd (1998) for further discussion.)

A further implication of the unconfoundedness assumption shown by Rosenbaum and Rubin (1983) is that $Y^0 \perp \Lambda \mid X$ is equivalent to

$$Y^0 \perp \Lambda \mid p(X),$$

where

$$p(x) \equiv P[\Lambda = 1|X = x]$$

is the treatment *propensity score*. The second term in Δ_1^* can then also be identified using the fact that

$$E(Y^0|\ \Lambda = 0, X) = E(Y^0|\ \Lambda = 0, p(X)) = E(Y^0|\ p(X)).$$

This creates significant opportunities for simplifying the analysis of effect by conditioning on the scalar $p(X)$ rather than on the vector X. See Rosenbaum and Rubin (1983) and Hirano, Imbens, and Ridden (2003) for more.

It is now straightforward to relate Rubin's unconfoundedness assumption to conditional exogeneity. We do this simply by making the identifications

$$\begin{aligned} Y^i &= c_i(\tilde{Z}, \ddot{Z}) \qquad i = 0, 1 \\ X &= (\tilde{Z}, \tilde{W}). \end{aligned}$$

Our next result establishes that conditional exogeneity, i.e. conditional independence given predictive proxies, implies unconfoundedness.

Proposition 4.1: *If* $\ddot{Z}_t \perp \Lambda_t \mid (\tilde{Z}_t, \tilde{W}_t)$, *then* $(c_0(\tilde{Z}_t, \ddot{Z}_t),\ c_1(\tilde{Z}_t, \ddot{Z}_t)) \perp \Lambda_t \mid (\tilde{Z}_t, \tilde{W}_t)$, *i.e. CIPP /conditional exogeneity implies unconfoundedness.*

One might ask whether CIPP is equivalent to unconfoundedness. Equivalence can indeed be shown if $\ddot{z}$ can be expressed as a function of y^0, y^1, and $\tilde{z}$

using $y^0 = c_0(\tilde{z}, \ddot{z})$, $y^1 = c_1(\tilde{z}, \ddot{z})$. Generally, $\ddot{z}$ will have to be of dimension 2 or less for this to be possible, and although this may hold in particular instances (e.g. additive separability of c_0 and c_1 between $\tilde{z}$ and $\ddot{z}$) this is not plausible for the general case.

What, then, does conditional exogeneity add, since it is generally stronger than unconfoundedness? The answer is that it provides insight that is particularly germane to economic applications. Proposition 4.1 establishes CIPP as a (perhaps *the*) natural primitive condition to ensure unconfoundedness, when the observable and unobservable determinants of the dependent variable are explicitly identified. This places a clear premium on economic theory in specifying the determining variables and on the need to carefully assess which determining variables are observable. Further, CIPP provides important insight into what constitute suitable covariates. Our discussion so far has suggested that suitable covariates are not only the observable determining factors $\tilde{Z}_t$ but also observable non-determining factors $\tilde{W}_t$ that serve as proxies for unobservable determining factors. In Sections 6 and 7 we examine the choice of auxiliary variables $\tilde{W}_t$ in depth. This will take us considerably beyond simply positing that some collection of covariates satisfies unconfoundedness, further highlighting the central role of economic theory in identifying the predictive proxies.

Now consider the consequences of conditional exogeneity. If $\breve{G}_0 = \breve{G}_1$,

$$\begin{aligned}\mu_{01} &= \int c_0(\tilde{z}, \ddot{z})\, d\breve{G}_1(\ddot{z}|x)\, d\breve{H}_1(x) \\ &= \int c_0(\tilde{z}, \ddot{z})\, d\breve{G}_0(\ddot{z}|x)\, d\breve{H}_1(x) \\ &= \int \breve{\mu}_0(x)\, d\breve{H}_1(x),\end{aligned}$$

where

$$\breve{\mu}_0(x) \equiv \int c_0(\tilde{z}, \ddot{z})\, d\breve{G}_0(\ddot{z}|x)$$

is the conditional expectation of $c_0(\tilde{Z}_t, \ddot{Z}_t)$ given $X_t \equiv (\tilde{Z}_t, \tilde{W}_t) = (\tilde{z}, \tilde{w}) \equiv x$.

It follows that one can straightforwardly obtain a consistent estimate of μ_{01} by estimating $\breve{\mu}_0$ consistently and then averaging over $\breve{H}_1$, the regime 1 distribution of $(\tilde{Z}_t, \tilde{W}_t)$. This is the approach taken by Hahn (1998) who estimates μ_{01} as

$$\hat{\mu}^H_{01} = T_1^{-1} \sum_{t=1}^{n} \Lambda_t \hat{\beta}_0(X_t)$$

where

$$\hat{\beta}_0(X_t) \equiv \frac{\hat{E}[(1-\Lambda_t)Y_t|X_t]}{1-\hat{E}[\Lambda_t|X_t]},$$

with $\hat{E}$ a nonparametric estimator of the indicated conditional expectation.

Hahn's estimator of the treatment effect on the treated ($\tilde{\gamma}$ in his notation) also makes use of an estimator for μ_1 constructed parallel to $\hat{\mu}_{01}^H$, namely

$$\hat{\mu}_1^H = T_1^{-1} \sum_{t=1}^{n} \Lambda_t \hat{\beta}_1(X_t),$$

where

$$\hat{\beta}_1(X_t) \equiv \frac{\hat{E}[\Lambda_t Y_t | X_t]}{\hat{E}[\Lambda_t | X_t]}.$$

Thus Hahn's estimator for the effect of interest is

$$\hat{\Delta}_H \equiv \hat{\mu}_1^H - \hat{\mu}_{01}^H.$$

This estimator requires nonparametric estimation of $E[Y_t|X_t]$, $E[\Lambda_t Y_t|X_t]$, and of the propensity score, $E(\Lambda_t|X_t)$. Hahn proposes the use of nonparametric series estimators (polynomials in X_t) for these conditional expectations and, in his Theorem 6, provides conditions ensuring that $\hat{\Delta}_H$ is consistent for Δ_1^*, asymptotically normal, and attains the semi-parametric efficiency bound.

Hahn's conditions include a random sampling assumption (i.i.d. observations) that is ideally suited for cross-section applications. For the natural experiments of interest here, the i.i.d. assumption is too strong. Nevertheless, weaker conditions permitting dependence (i.e. suitable martingale difference conditions, together with α- or β-mixing conditions on $\tilde{Z}_t$, $\ddot{Z}_t$, and $\tilde{W}_t$) should deliver analogous asymptotic normality results. Whether asymptotic efficiency condinues to hold is an interesting topic for further research.

Hirano, Imbens, and Ridder (2003) (HIR) note that the burden of nonparametric estimation posed by Hahn's estimator can be significantly reduced by exploiting the properties of the propensity score. Among the estimators they consider is an estimator of the average treatment effect on the treated that can be written

$$\hat{\Delta}_{HIR} = \hat{\mu}_1 - \hat{\mu}_{01}^{HIR},$$

where $\hat{\mu}_1 \equiv T_1^{-1} \sum_{t=1}^{n} \Lambda_t Y_t$ is the regime 1 sample mean as above, and

$$\hat{\mu}_{01}^{HIR} \equiv T_1^{-1} \sum_{t=1}^{n} (1-\Lambda_t) Y_t \ \hat{p}(X_t) \ / \ (1-\hat{p}(X_t)),$$

with $\hat{p}$ a nonparametric estimator of the propensity score. The HIR estimator eliminates the need to estimate $E[Y_t \mid X_t]$ and $E[\Lambda_t Y_t \mid X_t]$. HIR propose a logistic series estimator for constructing $\hat{p}$, and in their Theorem 5 provide conditions under which $\hat{\Delta}_{HIR}$ is consistent for Δ_1^*, asymptotically normal, and also attains the semiparametric efficiency bound. HIR's conditions also include an i.i.d. assumption, but, just as for Hahn's estimator, the extension of the asymptotic normality results to the dependent case should be straightforward.

An interesting feature of the HIR estimator is that $\hat{\mu}_{01}^{HIR}$ is not constructed by averaging over the observations of regime 1, but is instead a weighted average over regime 0. Heuristically, the effect is to approximate

$$\mu_{00}(h) \equiv \int \breve{\mu}_0(x) h(x)\, d\breve{H}_0(x)$$

with h chosen as $d\breve{H}_1 / d\breve{H}_0$, as then

$$\mu_{00}(d\breve{H}_1 / d\breve{H}_0) = \mu_{01}.$$

HIR's estimator approximates $d\breve{H}_1 / d\breve{H}_0$ as

$$\hat{h}(x) = (T_0/T_1)\hat{p}(x)/(1 - \hat{p}(x))$$

and then averages over the empirical distribution of $X_t \equiv (\tilde{Z}_t, \tilde{W}_t)$ in regime 0.

Thus, the researcher interested in estimating the effect of a natural experiment such as a cartel or a merger can do so consistently using $\hat{\Delta}_H$ or $\hat{\Delta}_{HIR}$ without having to satisfy the restrictive conditions ensuring the validity of the dummy variable approach. Specifically, there is no need to correctly specify models for c_0 and c_1, it is not necessary to achieve conditional independence of $\ddot{Z}_t$ and Λ_t given $\tilde{Z}_t$ ($\tilde{G}_0 = \tilde{G}_1$) and/or $H_0 = H_1$ or $m_0 = m_1$, and included variables need not be error free, but may be error-laden proxies for otherwise unobservable determining variables. Heuristically, the freedom from the need for correct specification is ensured by the nonparametric estimation of the conditional expectations. The remaining benefits are provided by the introduction of the auxiliary variables $\tilde{W}_t$ delivering predictive proxies $X_t \equiv (\tilde{Z}_t, \tilde{W}_t)$ satisfying CIPP and therefore unconfoundedness.

5 An Alternative Quasi-Nonparametric Estimator

Perhaps the leading reasons for the common application of the dummy variable approach to estimate the effects of interest here are its familiarity and its simple computation. The estimators $\hat{\Delta}_H$ and $\hat{\Delta}_{HIR}$ are less familiar and require a potentially challenging nonprametric estimation. In this section we propose a new "quasi-nonparametric" estimator for the effect of a natural experiment that requires only a straightforward dummy variable regression, thus achieving the advantages of familiarity and ease of computation, but that can also achieve an arbitrarily close approximation to the effect of interest, due to its flexibility.

Our approach is quasi-nonparametric in that its flexibility provides the basis for full nonparametric estimation, but for simplicity we do not pursue the technicalities to fully justify this here. Rather, we content ourselves with setting forth the basic method and establishing its properties under simpler conditions than are required for fully justifying our approach as nonparametric. This latter task may be taken up elsewhere. Our regularity conditions are sufficiently flexible as to deliver the consistency and asymptotic normality of our quasi-nonparametric estimator in either cross-section or time-series settings.

Our approach estimates the effect of interest as

$$\hat{\Delta}_1 = \hat{\mu}_1 - \hat{\mu}_{01},$$

where

$$\hat{\mu}_{01} = T^{-1} \sum_{t=1}^{n} \Lambda_t \widehat{\breve{\mu}}_0(X_t)$$

and $\widehat{\breve{\mu}}_0$ is an estimator of

$$\breve{\mu}_0(x) \equiv \int c_0(\tilde{z}, \ddot{z}) \, d\breve{G}_0(\ddot{z}|x).$$

Thus, our estimate of the effect of interest constitutes a predictive approach, in that it is constructed as the difference between the average actual outcome, $\hat{\mu}_1$, and the average outcome predicted for regime 1 under the retime 0 predictive relation.

To motivate our approach, we will proceed as if

$$\begin{aligned} \breve{\mu}_0(x) &= \sum_{j=0}^{q-1} \psi_j(x) b^*_{\psi,0,j} \\ &= x'_\psi b^*_{\psi,0}, \end{aligned}$$

where x_ψ is the $q \times 1$ vector with elements $\psi_0(x) = 1$, $\psi_j(x)$, ψ_j a given known measurable function, $j = 1, \ldots, q-1$; and $b^*_{\psi,0}$ is an unknown $q \times 1$ vector of coefficients to be estimated, having elements $b^*_{\psi,0,j}$, $j = 0, \ldots, q-1$. Suitable choice of q and $\{\psi_j\}$ yields a vast array of flexible functional forms, including the polynomial series proposed by Hahn (1998) and HIR, as well as Fourier series, artificial neural networks, and wavelets. These flexible forms can well approximate whatever the true function $\breve{\mu}_0$ might be. (See Hornik, Stinchcombe, and White (1989, 1990), White (1990), Gallant and White (1992), Stinchcombe and White (1998), Gencay, Selchuk, and Whitcher (2001), and White (2004).)

Estimation of $b^*_{\psi,0}$ is straightforward. An obvious consistent method is OLS regression of Y_t on $X_{\psi t} \equiv (1, \psi_1(X_t), ..., \psi_{q-1}(X_t))'$ using regime 0 observations:

$$\hat{\beta}_{\psi,0} \equiv \left(\sum_{t=1}^{n} (1 - \Lambda_t) X_{\psi t} X'_{\psi t} \right)^{-1} \sum_{t=1}^{n} (1 - \Lambda_t) X_{\psi t} Y_t.$$

This yields

$$\widehat{\breve{\mu}}_0(x) = x'_\psi \hat{\beta}_{\psi,0}$$

and, with $\hat{m}_{\psi,1} \equiv T_1^{-1} \sum_{t=1}^{n} \Lambda_t X_{\psi t}$,

$$\begin{aligned} \hat{\mu}_{01} &= T_1^{-1} \sum_{t=1}^{n} \Lambda_t X'_{\psi t} \hat{\beta}_{\psi,0} \\ &= \hat{m}'_{\psi,1} \hat{\beta}_{\psi,0}, \end{aligned}$$

Our quasi-nonparametric estimate of the efffect of interest is thus

$$\hat{\alpha}_\psi \equiv \hat{\mu}_1 - \hat{m}'_{\psi,1}\hat{\beta}_{\psi,0}.$$

By standard properties of OLS, $\hat{\alpha}_\psi$ can be obtained by applying OLS to

$$Y_t = \Lambda_t \alpha + X'_{\psi t}\beta + \Lambda_t \tilde{X}'_{\psi t}\gamma + \varepsilon_t, \qquad t = 1, 2, \ldots,$$

where $X_{\psi t} \equiv (1, \tilde{X}'_{\psi t})'$. Note the close resemblance of this model to the dummy variable model of Section 3. The only differences are: (1) we use regressors $X_{\psi t}$ (that include flexible transformations of X_t) as opposed to regressors $Z_t = (1, \tilde{Z}_t)$; and (2) we include $\Lambda_t \tilde{X}_{\psi t}$ in the regression.

Our first formal condition extends the DGP to accommodate $\tilde{W}_t$.

Assumption B.1 (Data Generating Process) *The observed data are generated from a realization of the sequence of random variables $(Y_t, \Lambda_t, \tilde{Z}_t, \ddot{Z}_t, \tilde{W}_t)$, $t = 1, 2, \ldots$, where $(\tilde{Z}_t, \ddot{Z}_t)$ stably isolates Λ_t for Y_t according to*

$$Y_t = c(\Lambda_t, \tilde{Z}_t, \ddot{Z}_t), \qquad t = 1, 2, \ldots$$

for some unknown measurable scalar-valued function c, where Λ_t is $\{0, 1\}$-valued, and Λ_t is fundamentally non-determining for $\tilde{W}_t$.

For $i = 0, 1$, define $\mathcal{T}_i \equiv \{t \in \mathbb{N} : \Lambda_t = i\}$, and assume further that for all $t \in \mathcal{T}_i$, $(\tilde{Z}_t, \ddot{Z}_t, \tilde{W}_t)$ has joint distribution $\breve{F}_i$, $(\tilde{Z}_t, \ddot{Z}_t)$ has joint distribution F_i, $(\tilde{Z}_t, \tilde{W}_t)$ has joint distribution $\breve{H}_i$, $\tilde{Z}_t$ has joint distribution H_i, $\ddot{Z}_t$ has joint distribution G_i, the conditional distribution of $\ddot{Z}_t$ given $(\tilde{Z}_t, \tilde{W}_t) = (\tilde{z}, \tilde{w})$ is *$\breve{G}_i(\cdot|\tilde{z}, \tilde{w})$, and the conditional distribution of $\ddot{Z}_t$ given $\tilde{Z}_t = \tilde{z}$ is $\tilde{G}_i(\cdot|\tilde{z})$.*

The variables $\tilde{W}_t$ represent observable non-determining variables that can act as useful proxies for the unobservable determining variables $\ddot{Z}_t$. Note that we permit $\tilde{W}_t$ to contain error-laden measures of determining variables that would conventionally be thought to create "errors in variables" problems.

Note the key requirement that Λ_t is fundamentally non-determining for $\tilde{W}_t$. Because $(\tilde{Z}_t, \ddot{Z}_t)$ stably isolates Λ_t, we have that Λ_t is fundamentally non-determining for $(\tilde{Z}_t, \tilde{W}_t)$ jointly. In the absence of this condition, indirect effects of the natural experiment can enter through $\tilde{W}_t$, whose regression coefficients will pick up effects properly due to the natural experiment, thereby contaminating the analysis.

Moment conditions for our flexible regressors are specified as follows.

Assumption B.2 (Finiteness of Moments) *For given $q \in \mathbb{N}$ and given known measurable scalar valued functions $\psi_0 = 1$, ψ_j, $j = 1, \ldots, q-1$, let $x_\psi \equiv (\psi_0(x), \ldots, \psi_{q-1}(x))'$, where $x \equiv (\tilde{z}, \tilde{w})$.*

(a) $0 < p_1 < 1$;

(b) (i) $M_{\psi,0} \equiv \int x_\psi x'_\psi d\breve{H}_0(x) < \infty, \quad \det M_{\psi,0} > 0$;

(ii) $m_{\psi,1} \equiv \int x_\psi d\breve{H}_1(x) < \infty$;

(c) (i) $L_{\psi,0} \equiv \int x_\psi c_0(\tilde{z}, \ddot{z})\, d\breve{G}_0(\ddot{z}|x)\, d\breve{H}_0(x) < \infty$;

(ii) $\mu_1 \equiv \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) < \infty$.

To state our next assumption, we define

$$\begin{aligned} \hat{M}_{\psi,0} &\equiv T_0^{-1} \sum_{t=1}^{n} (1 - \Lambda_t) X_{\psi t} X'_{\psi t} \\ \hat{L}_{\psi,0} &\equiv T_0^{-1} \sum_{t=1}^{n} (1 - \Lambda_t) X_{\psi t} Y_t. \end{aligned}$$

The law of large numbers requirement for our predictive approach is

Assumption B.3 (Laws of Large Numbers)

(a) $\hat{p}_1 \xrightarrow{p} p_1$;

(b) $\hat{M}_{\psi,0} \xrightarrow{p} M_{\psi,0}$ *and* $\hat{m}_{\psi,1} \xrightarrow{p} m_{\psi,1}$;

(c) $\hat{L}_{\psi,0} \xrightarrow{p} L_{\psi,0}$ *and* $\hat{\mu}_1 \xrightarrow{p} \mu_1$.

As before, this condition is mild and thus generally plausible. We now have

Proposition 5.1 *Given B.1-B.3, $\hat{\alpha}_\psi \xrightarrow{p} \alpha^*_\psi$ and $\hat{\beta}_{\psi,0} \to \beta^*_{\psi,0}$, where*

$$\begin{aligned} \alpha^*_\psi &\equiv \mu_1 - m'_{\psi,1}\beta^*_{\psi,0} \\ \beta^*_{\psi,0} &\equiv M^{-1}_{\psi,0} L_{\psi,0}. \end{aligned}$$

Proposition 5.2 *Suppose Assumptions B.1 and B.2 hold and that in addition $\mu_{01} \equiv \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) < \infty$. Then*

$$\begin{aligned} \alpha^*_\psi - \Delta^*_1 &= \int (\breve{\mu}_{01}(x) - \breve{\mu}_0(x))\; d\breve{H}_1(x) \\ &\quad + \int (\breve{\mu}_0(x) - x'_\psi \beta^*_{\psi,0})\; d\breve{H}_1(x), \end{aligned}$$

where $\breve{\mu}_{01}(x) \equiv \int c_0(\tilde{z}, \ddot{z})\, d\breve{G}_1(\ddot{z}|x)$.

Observe that this result stops short of establishing that the causal discrepancy $\alpha^*_\psi - \Delta^*_1$ vanishes. Instead it decomposes it into two pieces, the first component depending on $\breve{\mu}_{01} - \breve{\mu}_0$ (hence $\breve{G}_1 - \breve{G}_0$) and the second depending on the error $\breve{\mu}_0(x) - x'_\psi \beta^*_{\psi,0}$ in approximating the optimal predictor $\breve{\mu}_0$. By suitably choosing proxies X_t so that $\breve{G}_0 = \breve{G}_1$, the economist can make $\breve{\mu}_{01} - \breve{\mu}_0$ vanish. By suitably choosing q and $\{\psi_j\}$, the economist can make the approximation error negligible. Together, these choices ensure that the causal discrepancy (essentially) vanishes. Stated formally, we have

Corollary 5.3 *Suppose Assumptions B.1 and B.2 hold and that $\mu_{01} < \infty$.*

(i) *Suppose further that $\breve{\mu}_0(x) = x_\psi' b_{\psi,0}^*$ for some unknown finite vector $b_{\psi,0}^*$. Then $\beta_{\psi,0}^* = b_{\psi,0}^*$.*

(ii) *If in addition $\breve{G}_0 = \breve{G}_1$ a.s. $- \breve{H}_1$, then $\breve{\mu}_0 = \breve{\mu}_{01}$ a.s. $- \breve{H}_1$ and*

$$\alpha_\psi^* = \Delta_1^*.$$

Thus our alternative approach can yield a consistent estimate of the average effect of the natural experiment under plausible conditions. Success is ensured by satisfying two key requirements: First, the researcher must achieve an accurate approximation to the conditional expectation $\breve{\mu}_0$. This requires some care and thought, but it is an attainable goal, especially given the availability of flexible modeling tools like PcGets (e.g., Hendry and Krolzig (2001) and Campos, Hendry, and Krolzig (2003)), RETINA (Perez-Amaral, Gallo, and White, 2003), and QuickNet (White, 2004). Second, the researcher must choose X_t so that $\breve{G}_0 = \breve{G}_1$. It is here that statistical reasoning leaves off and where economic understanding assumes the leading role, as we discuss in detail in Sections 6 and 7.

Our estimator is quasi-nonparametric in that although we have formally treated only the parametric case here, a fully nonparametric estimator can be attained by letting $q \to \infty$ as $n \to \infty$ at the appropriate rate with suitable choice of $\{\psi_j\}$. Results of Chen and White (1999) may be useful in establishing asymptotic properties.

A main focus of Hahn (1998) and Hirano, Imbens, and Ridder (2003) is to establish the asymptotic normality of their estimators and their attainment of the semiparametric efficiency bound. Our next results permit a comparison of $\hat{\alpha}_\psi$ with $\hat{\Delta}_H$ and $\hat{\Delta}_{HIR}$. For this, we provide conditions ensuring the asymptotic normality of $\hat{\alpha}_\psi$ and examine the asymptotic variance.

Our first condition to this end is a moment condition analogous to B.2:

Assumption B.4 (Finiteness of Moments): *For q and x_ψ as in B.2,*

(a) $M_{\psi,1} \equiv \int x_\psi x_\psi' d\breve{H}_1(x) < \infty$, det $M_{\psi,1} > 0$;

(b) $L_{\psi,1} \equiv \int x_\psi c_1(\tilde{z}, \ddot{z})\ d\breve{G}_1(\ddot{z}|x)\ d\breve{H}_1(x) < \infty$.

This condition ensures the existence of

$$\beta_{\psi,1}^* \equiv M_{\psi,1}^{-1} L_{\psi,1},$$

which is the vector of regression coefficients delivering the *m.s.e.*-optimal approximation to $\breve{\mu}_1(x)$, which has the form

$$\mu_1^*(x_\psi) \equiv x_\psi' \beta_{\psi,1}^*.$$

We analogously define the *m.s.e.*-optimal approximation to $\breve{\mu}_0(x)$,

$$\mu_0^*(x_\psi) \equiv x_\psi' \beta_{\psi,0}^*.$$

The *m.s.e.*-optimality property ensures that

$$\mu_1 = \int \mu_1^*(x_\psi) \; d\breve{H}_1(x)$$

so that

$$\int (\mu_1^*(x_\psi) - \mu_0^*(x_\psi)) \; d\breve{H}_1(x) = \mu_1 - m_{\psi,1}' \beta_{\psi,0}^* = \alpha_\psi^*.$$

We also define the regression errors

$$\begin{aligned} \varepsilon_{t0} &\equiv Y_t - X_{\psi t}' \beta_{\psi,0}^*, \\ \varepsilon_{t1} &\equiv Y_t - X_{\psi t}' \beta_{\psi,1}^*, \end{aligned}$$

and we note that the *m.s.e.*-optimality properties of $\beta_{\psi,0}^*$ and $\beta_{\psi,1}^*$ ensure that ε_{t0} is uncorrelated with $X_{\psi t}$ in regime 0 and ε_{t1} is uncorrelated with $X_{\psi t}$ in regime 1. Our next condition can now be stated, with $p_0 \equiv (1 - p_1)$.

Assumption B.5 (Central Limit Conditions)

(a) $n^{-1/2} \sum_{t=1}^n (1 - \Lambda_t) X_{\psi t} \varepsilon_{t0} = O_p(1)$;

(b) $n^{-1/2} \sum_{t=1}^n \Lambda_t [\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)] = O_p(1)$;

(c) $n^{-1/2} \sum_{t=1}^n \xi_t \xrightarrow{d} N(0, \sigma_\xi^2)$, *where* $0 < \sigma_\xi^2 \equiv var(n^{-1/2} \sum_{t=1}^n \xi_t) < \infty$,

$$\xi_t \equiv p_1^{-1} \Lambda_t [\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)] - p_0^{-1} m_{\psi,1}' M_{\psi,0}^{-1} (1 - \Lambda_t) X_{\psi t} \varepsilon_{t0}.$$

Assumptions B.1, B.2(a), B.2(b.i), B.2 (c.i), B.3 and B.5(a) ensure that

$$\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*) = p_0^{-1} M_{\psi,0}^{-1} n^{-1/2} \sum_{t=1}^n (1 - \Lambda_t) X_{\psi t} \varepsilon_{t0} + o_p(1).$$

The summands on right hand side above have mean zero, as ε_{t0} is uncorrelated with $X_{\psi t}$ in regime 0, so B.5(a) holds given that $\{(1 - \Lambda_t) X_{\psi t} \varepsilon_{t0}\}$ obeys a central limit theorem (CLT). Primitive conditions for this applicable either to i.i.d. or time-series settings are given by White (2001, ch.5) For example, certain α-mixing conditions suffice in the time-series case. When such a CLT holds, we also have that $\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*)$ is asymptotically normal.

Similarly, a CLT for $\{\Lambda_t[\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*]\}$ suffices for B.5(b). Note that the properties of ε_{t1}, μ_1^*, μ_0^* and α_ψ^* ensure that $\Lambda_t[\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - a_\psi^*)]$ has mean zero under regime 1.

Assumption B.5(c) delivers our asymptotic normality result, as we prove

$$\sqrt{n}(\hat{\alpha}_\psi - \alpha_\psi^*) = n^{-1/2} \sum_{t=1}^{n} \xi_t + o_p(1).$$

Theorem 5.4: *Suppose conditions B.1-B.5 hold. Then*

$$\sqrt{n}(\hat{\alpha}_\psi - \alpha_\psi^*) \xrightarrow{d} N(0, \sigma_\xi^2).$$

Thus, our quasi-nonparametric estimator of the effect of interest is not only consistent for α_ψ^* (approximating Δ_1^*, and equal to it under the conditions of Corollary 5.3), but it is also asymptotically normal with asymptotic variance σ_ξ^2. This variance can be consistently estimated using HAC methods, e.g. under conditions ensuring the consistent estimation of the variance-covariance matrix for the OLS estimator of the linear regression (5.1) (see White 2001, ch. 6 or Gonçalves and White, 2005), so we shall not go into specifics here.

Instead, we compare σ_ξ^2 with the asymptotic variance of the Hahn and HIR. As HIR show,

$$\begin{aligned} avar(\hat{\Delta}_H) &= avar(\hat{\Delta}_{HIR}) \\ &= T_1^{-1} \int \sigma_1^2(x)\ d\breve{H}_1(x) \\ &\quad + T_1^{-1} \int (\breve{\mu}_1(x) - \breve{\mu}_0(x) - \Delta_1^*)^2\ d\breve{H}_1(x) \\ &\quad + T_0^{-1} \int \sigma_0^2(x)(d\breve{H}_1(x)/\ d\breve{H}_0(x))^2 d\breve{H}_0(x) \end{aligned}$$

where we have re-scaled á la Froehlich (2004) to facilitate comparisons and interpretations (cf Froehlich, 2004, eq (6)). In the expression above, we define

$$\begin{aligned} \sigma_0^2(x) &\equiv \int [c_0(\tilde{z}, \ddot{z}) - \breve{\mu}_0(x)]^2 d\breve{G}_0(\ddot{z}|x) \\ \sigma_1^2(x) &\equiv \int [c_1(\tilde{z}, \ddot{z}) - \breve{\mu}_1(x)]^2 d\breve{G}_1(\ddot{z}|x). \end{aligned}$$

Direct comparison with σ_ξ^2 is possible only when the summands $\{\xi_t\}$ are uncorrelated, as otherwise additional terms enter that capture possible time-series dependence. Thus, suppose that $\{\xi_t\}$ is i.i.d. Analogous rescaling then gives

$$\begin{aligned} avar(\hat{\alpha}_\psi) &= T_1^{-1} \int \sigma_1^{*2}(x)\ d\breve{H}_1(x) \\ &\quad + T_1^{-1} \int (\mu_1^*(x_\psi) - \mu_0^*(x_\psi) - \alpha_\psi^*)^2 d\breve{H}_1(x) \\ &\quad + T_0^{-1} \int \sigma_0^{*2}(x) g(x)^2 d\breve{H}_0(x) \end{aligned}$$

where

$$\sigma_0^{*2}(x) \equiv \int (c_0(\tilde{z},\ddot{z}) - \mu_0^*(x_\psi))^2 d\breve{G}_0(\ddot{z}|x)$$
$$\sigma_1^{*2}(x) \equiv \int (c_1(\tilde{z},\ddot{z}) - \mu_1^*(x_\psi))^2 d\breve{G}_1(\ddot{z}|x)$$

are the conditional variances of ε_{t0} and ε_{t1} in regimes 0 and 1 respectively, and

$$g(x) \equiv x_\psi' M_{\psi,0}^{-1} m_{\psi,1}.$$

The expressions for $avar(\hat{\Delta}_{HIR}) = avar(\hat{\Delta}_H)$ and for $avar(\hat{\alpha}_\psi)$ clearly have a similar structure, with the primary diffrences arising because the Hahn and HIR estimators are fully nonparametric, whereas $\hat{\alpha}_\psi$ is only quasi-nonparametric. To make the comparison more direct, suppose for simplicity that we have correct specification as in Corollary 5.3, so that $\mu_0^*(x_\psi) = \breve{\mu}_0(x)$ and similarly that $\mu_1^*(x_\psi) = \breve{\mu}_1(x)$. Then $\alpha_\psi^* = \Delta_1^*$, $\sigma_0^* = \sigma_0$, and $\sigma_1^* = \sigma_1$. We then have

$$avar(\hat{\alpha}_\psi) - avar(\hat{\Delta}_H)$$
$$= T_0^{-1} \int \sigma_0^2(x)[g(x)^2 - \frac{d\breve{H}_1(x)^2}{d\breve{H}_0(x)^2}]\ d\breve{H}_0(x).$$

It is not obvious from this expression that the two estimators bear an unambiguous efficiency relation to one another. If, for example, we have correct specification and the regression errors are conditionally normal and homoskedastic, then $\hat{\alpha}_\psi$ is the maximum likelihood estimator (MLE), which is the asymptotically efficient parametric estimator. Accordingly, it should be at least as efficient asymptotically as $\hat{\Delta}_H$ or $\hat{\Delta}_{HIR}$. When $\hat{\alpha}_\psi$ is not the MLE, then a sufficient condition for $\hat{\alpha}_\psi$ to achieve efficiency equal to that of $\hat{\Delta}_H$ or $\hat{\Delta}_{HIR}$ is

$$|g(x)| \equiv d\breve{H}_1(x)/d\breve{H}_0(x).$$

The density ratio $d\breve{H}_1/\ d\breve{H}_0$ is related to the propensity score p by

$$d\breve{H}_1(x)/\ d\breve{H}_0(x) = \frac{(1-p_1)}{p_1}\frac{p(x)}{1-p(x)}.$$

A sufficient condition for $\hat{\alpha}_\psi$ to attain the semi-parametric bound is thus

$$p(x) = |g(x)|\ /\ (|g(x)| + (1-p_1)/p_1).$$

This condition is not one that might be expected to hold generally, so it is not possible to assert that $\hat{\alpha}_\psi$ has any necessary efficiency properties in the i.i.d. case. Nevertheless, the simplicity with which $\hat{\alpha}_\psi$ and a consistent estimate of its standard error can be computed using standard regression packages makes it an appealing option even in cross-section applications.

For time series applications, the properties of $\hat{\Delta}_H$ and $\hat{\Delta}_{HIR}$ are unknown, and it is not clear whether they may still achieve the semiparametric efficiency bound, as not only is this bound presently unknown, but the asymptotic variances of $\hat{\Delta}_H$ and $\hat{\Delta}_{HIR}$ are also unknown in the time series context. It is of clear interest to analyze $\hat{\Delta}_H$ and $\hat{\Delta}_{HIR}$ in time-series settings. In the meantime, however, $\hat{\alpha}_\psi$ provides a convenient estimataor with known properties in both time-series and cross-section applications.

6 Justifying Conditional Exogeneity

As we have seen, some form of conditional independence is necessary in order to consistently estimate the effect of interest. This assumption must be justified in each particular application.

We note first that if the unconditional distribution of all the determining variables is identical between regimes, then it suffices to take $\mathbf{Z}_t = \ddot{Z}_t$, treating all determining variables as unobservable. No predictive proxies are needed, and we have $G = \breve{G}_0 = \breve{G}_1$. This assumption is readily ensured by randomization, but economists cannot assign treatment in natural experiments. If even one component of $\mathbf{Z}_t = \ddot{Z}_t$ has a distribution differing across regimes, then conditional exogeneity fails.

It is therefore critical to account for components of $\mathbf{Z}_t$ whose distributions differ across regimes. If such variables are precisely observable, then they can be included in $\tilde{Z}_t$. If the observable determining variables $\tilde{Z}_t$ are the only predictive proxies, then CIPP requires $\tilde{G}_0 = \tilde{G}_1$. Again, this assumption is readily ensured by randomization, but economists typically cannot achieve this. Absent further justification, the economist cannot plausibly treat $\tilde{Z}_t$ alone as predictive proxies ensuring CIPP.

How, then, can one go about selecting additional predictive proxies? The ideal predictive proxy for any element of $\ddot{Z}_t$ is itself, but $\ddot{Z}_t$ is by convention unobservable. Alternatively and more generally, a set of perfect proxies is provided by any information-preserving transform of $\ddot{Z}_t$, say $\tilde{W}_t = \zeta(\ddot{Z}_t)$, where ζ is information-preserving in the sense that the $\sigma-$field generated by $\ddot{Z}_t$, denoted $\sigma(\ddot{Z}_t)$, coincides with the $\sigma-$field generated by $\tilde{W}_t$, $\sigma(\tilde{W}_t)$. For example, even if $\ddot{Z}_t$ is not observable, it would suffice to observe a non-singular linear (or nonlinear) transformation of $\ddot{Z}_t$, even if the transformation is unknown.

Another possibility is to exploit the determining chain of Section 2 The next result provides conditions linking functional dependence and statistical independence in such a way as to expose potential predictive proxies.

Proposition 6.1 *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, and let $\mathcal{V} : \Omega \to \mathbb{R}^k$, $\mathcal{Z} : \Omega \to \mathbb{R}^\ell$, $k, \ell \in \mathbb{N}$ be random vectors. Let $f : \mathbb{R}^k \times \mathbb{R}^\ell \to \mathbb{R}^m$ be a measurable function so that $\mathcal{Y} \equiv f(\mathcal{V}, \mathcal{Z})$ is a random $m \times 1$ vector. Then $\mathcal{Y} \perp \mathcal{V} \mid \mathcal{Z}$ if and only if there is a measurable function $g : \mathbb{R}^\ell \to \mathbb{R}^m$ such that $\mathcal{Y} = g(\mathcal{Z})$ with probability* 1.

Suppose $\mathbf{Z}_t \equiv (\tilde{Z}_t, \ddot{Z}_t)$ isolates Λ_t and there exists $\mathbf{Z}_t^{(1)}$, say, also isolating

Λ_t, such that

$$\mathbf{Z}_t = \zeta(\mathbf{Z}_t^{(1)}).$$

Applying Proposition 6.1 with $\mathcal{Y} = \mathbf{Z}_t$, $\mathcal{Z} = \mathbf{Z}_t^{(1)}$ and $\mathcal{V} = \Lambda_t$ we have

$$\mathbf{Z}_t \perp \Lambda_t \,|\, \mathbf{Z}_t^{(1)},$$

so that in particular $\ddot{Z}_t \perp \Lambda_t \,|\, \mathbf{Z}_t^{(1)}$. As $\mathbf{Z}_t^{(1)}$ is non-determined by Λ_t, it follows that $\mathbf{Z}_t^{(1)}$ provides admissible predictive proxies. Any observable element of $\mathbf{Z}_t^{(1)}$ can be included in $\tilde{W}_t$; moreover, one can continue, sequentially, along the determining chain, appending observable elements of $\mathbf{Z}_t^{(j)}$ to $\tilde{W}_t$.

In practice, however, it may not be possible to accurately observe sufficient determining variables at whatever level of the determining chain to proxy for those components of $\mathbf{Z}_t$ whose conditional distributions differ across regimes. Instead, one may only have available error-laden proxies. To examine the content of Proposition 5.2 in a realistic setting that clearly exposes the key elements, suppose that none of the determining variables are observable without error, so that $\tilde{Z}_t$ is null. For notational convenience in what follows, we write $\mathbf{Z}_t = (1, \ddot{Z}_t)'$ and $\mathbf{z} = (1, \ddot{z})'$, and we assume that the regime 0 determining reduced form is given by

$$c_0(\tilde{z}, \ddot{z}) = \mathbf{z}' b_0^*.$$

Little generality is lost in writing the determining relationship in this way as information-preserving transformations applied to determining variables can yield neural-network-like structures analogous to those for $\breve{\mu}_0$ of Section 5.

We proceed by obtaining expressions for $\breve{\mu}_0$ and $\breve{\mu}_{01}$. Thus, suppose that we have predictive proxies $\tilde{W}_t$ fundamentally non-determined by Λ_t, and construct $X_{\psi t}$ as in Section 4: $X_{\psi t} \equiv (1, \psi_1(\tilde{W}_t), \ldots, \psi_{q-1}(\tilde{W}_t))'$, where we drop explicit reference to the absent $\tilde{Z}_t$.

Then

$$\beta^*_{\psi,0} = M^{-1}_{\psi,0}, L_{\psi,0},$$

where

$$\begin{aligned} L_{\psi,0} &\equiv \int x_\psi c_0(\tilde{z}, \ddot{z})\, d\breve{G}_0(\ddot{z} \,|\, x)\, d\breve{H}_0(x) \\ &= \int x_\psi \mathbf{z}' b_0^*\, d\breve{G}_0(\ddot{z} \,|\, \tilde{w})\, d\breve{H}_0(\tilde{w}) \\ &= K_{\psi,0}\, b_0^*, \end{aligned}$$

where in the second line we drop explicit reference to $\tilde{z}$, and we define

$$K_{\psi,0} \equiv \int x_\psi \mathbf{z}'\, d\breve{G}_0(\ddot{z} \,|\, \tilde{w})\, d\breve{H}_0(\tilde{w}).$$

Next, let

$$a_0^* \equiv M^{-1}_{\psi,0} K_{\psi,0}.$$

This matrix of coefficients ensures that

$$\mathbf{Z}_t' = X_{\psi t}' a_0^* + \varepsilon_t' \qquad t \in \mathcal{T}_0,$$

and $E(X_{\psi t}\varepsilon_t') = 0$ for $t \in \mathcal{T}_0$. This expression tells us that $X_{\psi t}' a_0^*$ gives an *m.s.e.*-optimal prediction for $\mathbf{Z}_t'$ in regime 0. It follows that

$$\begin{aligned} \beta_{\psi,0}^* &= M_{\psi,0}^{-1} K_{\psi,0}\, b_0^* \\ &= M_{\psi,0}^{-1} M_{\psi,0}\, a_0^*\, b_0^* \\ &= a_0^* b_0^*. \end{aligned}$$

With c_0 as specified, we have

$$\begin{aligned} \breve{\mu}_0(\tilde{w}) &= x_\psi' \beta_{\psi,0}^* \\ &= x_\psi'\, a_0^* b_0^*. \end{aligned}$$

Before proceeding, it is important to note that the regression coefficients $\beta_{\psi,0}^* = a_0^* b_0^*$ obtained by regressing Y_t on $X_{\psi t}$ are a blend of coefficients b_0^* that embody the true effects of the underlying determinants $\mathbf{Z}_t$ on Y_t and coefficients a_0^* that embody only predictive relationships between $\mathbf{Z}_t$ and proxies $X_{\psi t}$. It follows that the coefficients $\beta_{\psi,0}^*$ cannot be interpreted as providing any reliable information about the "effects" of the associated regressors $X_{\psi t}$. Put somewhat differently, one should not expect the coefficients $\beta_{\psi,0}^*$ to make economic sense, that is, to have signs and magnitudes that one might expect if one had performed regression using a correctly specified model. Thus, the coefficients $\beta_{\psi,0}^*$ enable the model *to account for* the effects of the proxied unobservable determinants $\mathbf{Z}_t$, rather than "controlling for" the effects of the included regressors $X_{\psi t}$.

Proceeding to $\breve{\mu}_{01}$, we have

$$\breve{\mu}_{01}(\tilde{w}) \equiv \int \mathbf{z}'\, b_0^*\, d\breve{G}_1(\ddot{z} \mid \tilde{w}).$$

Define

$$\begin{aligned} a_1^* &\equiv M_{\psi,1}^{-1} K_{\psi,1} \\ K_{\psi,1} &\equiv \int x_\psi \mathbf{z}'\, d\breve{G}_1(\ddot{z} \mid \tilde{w})\, d\breve{H}_1(\tilde{w}), \end{aligned}$$

so that

$$\mathbf{Z}_t' = X_{\psi t}' a_1^* + \varepsilon_t' \qquad t \in \mathcal{T}_1,$$

and $E(X_{\psi t}\varepsilon_t') = 0$ for $t \in \mathcal{T}_1$. That is, $X_{\psi t}' a_1^*$ gives an *m.s.e*-optimal prediction for $\mathbf{Z}_t'$ in regime 1. The properties of least squares approximation ensure

$$m_{\psi,1}' a_1^* = \int \mathbf{z}'\, d\breve{G}_1(\ddot{z} \mid \tilde{w})\, d\breve{H}_1(\tilde{w}).$$

Substituting these results into the causal discrepancy of Proposition 5.2 gives

$$\begin{aligned}\alpha_{\psi}^{*}-\Delta_{1}^{*} &= \int (\breve{\mu}_{01}(\tilde{w})-\breve{\mu}_{0}(\tilde{w}))\, d\breve{H}_{1}(\tilde{w}) \\ &\quad + \int (\breve{\mu}_{0}(\tilde{w})-x_{\psi}'\beta_{\psi,0}^{*})\, d\breve{H}_{1}(\tilde{w}) \\ &= \int \breve{\mu}_{01}(\tilde{w})\, d\breve{H}_{1}(\tilde{w}) - \int \breve{\mu}_{0}(\tilde{w})\, d\breve{H}_{1}(\tilde{w}) \\ &= m_{\psi,1}' a_{1}^{*} b_{0}^{*} - \int x_{\psi}' a_{0}^{*} b_{0}^{*}\ d\breve{H}_{1}(\tilde{w}) \\ &= m_{\psi,1}'(a_{1}^{*}-a_{0}^{*}) b_{0}^{*}.\end{aligned}$$

Thus, a sufficient condition for the causal discrepancy to vanish is

$$a_{1}^{*}=a_{0}^{*}.$$

If $\dim(\mathbf{Z}_t) \geq \dim(X_{\psi,t})$, then $a_1^* - a_0^*$ generally will have full column rank, so that $a_1^* = a_0^*$ is also a necessary condition for the causal discrepancy to vanish.

The key condition $a_1^* = a_0^*$ succinctly expresses the requirement that the predictive relation between the unobservable determinants $\mathbf{Z}_t$ and the predictive proxies $X_{\psi t}$ is stable across regimes, that is ,

$$\mathbf{Z}_t' = X_{\psi t}' a^* + \varepsilon_t' \qquad t \in \mathcal{T}_0 \cup \mathcal{T}_1, \tag{6.1}$$

where $a^* = a_0^* = a_1^*$, and $E(X_{\psi t}\varepsilon_t') = 0$ for all t. This predictive stability is the essence of CIPP. Indeed, CIPP implies $a_1^* = a_0^*$ for any transformation ψ.

In the absence of the experimental control necessary to assign treatment randomly, it is not possible to affirmatively verify conditional exogeneity. Nevertheless, conditional exogeneity is subject to refutation, and this can be approached on both theoretical and empirical grounds. In such circumstances, we view justifying conditional exogeneity as the process of identifying and removing possible grounds for refutation. In the next section we discuss statistical (thus empirical) methods for attempting to refute this assumption. The remainder of this section discusses *a priori* economic considerations that bear on possible refutation of conditional exogeneity.

We proceed by considering how to obtain a properly justified set of predictive proxies. It is simplest to start from the situation in which the natural experiment is viewed as the only observable determinant of the variable of interest and all other determinants are unobservable. Thus we have

$$Y_t = c(\Lambda_t, \ddot{Z}_t^0),$$

where $\ddot{Z}_t^0$ denotes the initial set of unobservable determinants that, as required by B.1, are neither directly nor indirectly determined by Λ_t. At this stage there are no predictive proxies; equivalently, we can say that the only predictive proxy is the constant unity.

We now consider whether to augment the predictive proxies to include observables. We must do so if there is any evidence that the distribution of

$\ddot{Z}_t^0$ differs between the two regimes, as otherwise CIPP fails. Although $\ddot{Z}_t^0$ is unobservable, indirect evidence of distributional shifts is often available from observable proxies for the elements of $\ddot{Z}_t^0$. In Section 7, we discuss tests based on this fact. For now, we proceed assuming knowledge *a priori* of such a shift or the possibility of such a shift.

Taking the cartel example for concreteness, suppose that the distribution of the price of an important raw material differs between regimes for reasons unrelated to the cartel. For concreteness, suppose the average price increases. Then we will observe an increase in average price of the cartelized product, but in the absence of any predictive proxy besides the constant, there is no way to separate out the effect of the cartel from the unrelated effect of the cost increase. That is, conditional exogeneity fails as a result of the unaccounted for change in the distribution of the raw material price.

Although the actual raw material price may not be available (for example, because transactions are conducted in a complex way and because accurate records of these complex transactions are not cost effective for the firm to keep), the economist may have access to monthly government or trade organization statistics that report some index (*i.e.*, market summary measure) of the price of the commodity. These monthly price data are almost certainly not the ideal raw material prices driving the firm's decisions, but they are an appealing candidate as a (purely) predictive proxy. Nevertheless, just as we require that the ideal raw material price must be unaffected by the cartel, so too must its proxy be unaffected by the cartel. In particular, the predictive relation between the actual price (that is, the ideal component(s) of $\ddot{Z}_t$) and the commodity price index must be relatively stable across regimes. This is often quite plausible, as the construction of commodity price indexes typically involves computing weighted averages of prices surveyed in some more or less standard way. As long as both the underlying surveyed prices and the weights used in constructing the weighted average are not appreciably impacted by the cartel, then we are justified in adding the commodity price index to our set of (purely) predictive proxies, now designated $\tilde{W}_t^1$. For concreteness, we assume that our first predictive proxy goes into $\tilde{W}_t^1$, leaving $\tilde{Z}_t$ empty for now.

The next step is to again ask whether there is any evidence that components of $\ddot{Z}_t$ have distributions (conditional now on $\tilde{W}_t^1$) that differ between regimes of the natural experiment. If not, we are done. If so, conditional exogeneity is in question, and further steps are required to resolve the challenge posed by these other unaccounted-for factors.

Continuing our cartel example, suppose that there is evidence of a shift in demand across regimes, occurring for reasons unrelated to the cartel. For example, suppose the cartelized product is one that is widely used by a broad array of purchasers ("downstream buyers"), sufficiently so that fluctuations in aggregate demand facing the downstream buyers for their products causes significant fluctuations in demand for the cartelized product.

In order to properly belong to $\ddot{Z}_t$, these demand shifters must be unaffected either directly or indirectly by the operation of the cartel, and we proceed under this assumption. This can be justified if the cartelized industry is of modest size relative to the economy as a whole. (If this assumption fails, Assumption B.1 is called into question.) As it is typically quite difficult to measure the ideal demand shifter components of $\ddot{Z}_t$, it will be necessary to resort to (purely) predictive proxies for these demand shifters.

To illustrate further challenges to conditional exogeneity that can arise, consider using as predictive proxies indexes of industrial production for industry segments representing the downstream buyers. In particular, the breadth of the industrial production index can play an important role.

Specifically, if the industrial production (IP) index is for a sector that contains the cartel, and if the cartel achieves its aims by restricting production, then Assumption B.1 is violated, as the IP index is being driven by the cartel activity. To the extent that the cartel is only a small part of the sector covered by the IP index, then the problem is mitigated. Better, however, would be to remove the cartel industry component from that sector's IP index.

Alternatively, if the IP index is for a sector in which the cartel's products are a critical input, so that a quantity restriction by the cartel curtails production in this sector, then once again the IP index for that sector is driven by the cartel activity and Assumption B.1 is in question. In this case one must seek predictive proxies for product demand for the downstream buyer's products.

Once one has located valid predictive proxies of this sort, they may be appended to $\tilde{W}_t^1$ to create $\tilde{W}_t^2$.

Proceeding in this way, one arrives finally at some collection of predictive proxies $X_t \equiv (\tilde{Z}_t, \tilde{W}_t)$ that have been justified to the extent that (a) elements of $\tilde{Z}_t$ are accurately measured determinants of Y_t not themselves determined by Λ_t; and (b) elements of $\tilde{W}_t$ are proxies for unobservable determinants $\ddot{Z}_t$ of Y_t, where $\ddot{Z}_t$ and $\tilde{W}_t$ are not determined by Λ_t, and the predictive relation between $\ddot{Z}_t$ and X_t is not impacted by Λ_t.

As is evident from Corollary 5.2, the impact of the violation of CIPP is a matter of degree. The more important the unobserved factors ($\ddot{z}$) and the greater the distributional shift $\breve{G}_1 - \breve{G}_0$, the greater is $\alpha_\psi^* - \Delta_1^*$, the causal discrepancy. Thus, the greatest care must be directed toward obtaining proxies for the most important determinants of the affected variable of interest and toward those for which the distributional shift $\breve{G}_1 - \breve{G}_0$ is potentially greatest. Neglect of minor determinants or of determinants for which $\breve{G}_1 - \breve{G}_0$ is small will have only minor impact on $\alpha_\psi^* - \Delta_1^*$.

We emphasize that this process does not constitute a verification of CIPP, but rather a justification, in that by using economic theory and the economist's knowledge of how economic data are constructed, one can avoid obvious (and perhaps not so obvious) opportunities for the refutation of conditional exogeneity and the violation of the requirement that Λ_t must be fundamentally non-determining for $(\tilde{Z}_t, \tilde{W}_t)$.

Yet another means of avoiding refutation of conditional exogeneity, especially in controversial settings, is to make use of the underlying factors and/or predictive proxies identified by parties adverse to one's own position that either explicitly or implicitly have been treated as exogenous to whatever degree. If one accounts for all legitimate factors articulated by one's rivals and nevertheless obtains statistically valid evidence in support of one's own position, then the onus is on the rivals to demonstrate why their original factors were inadequate to support their position — not an enviable position.

Note the importance of the qualifier "legitimate" in the previous sentence. Predictive proxies that directly or indirectly contain the effects of cartel operation are not legitimate, as this violates B.1. All variables determinatively linked to variables controlled by cartel members, such as prices, quantities, or inventories, among other things, are illegitimate for this reason. Also not legitimate are predictive proxies that do not have well-justified economic links to the elements of $\ddot{z}$. Economic time series are sufficiently numerous that one can often find a data series with just the right pattern to "explain" the effect of the natural experiment of interest. A strong antidote to curve-fitting of this sort, surreptitious or otherwise, is economic theory. The absence of a viable economic justification for including a particular proxy is strong grounds for its exclusion, given how easily such variables can mask the true effect.

Economic theory thus plays a decisive role in isolating and analyzing the effects of a natural experiment, and it is useful to itemize the various aspects of this role. Specifically, economic theory identifies: (1) the underlying determining factors $\mathbf{z}$ not caused by the operation of the natural experiment; (2) variables that are jointly determined by the operation of the natural experiment and that are therefore not legitimate components of $\mathbf{z}$; (3) variables that may or may not therefore form the basis for legitimate predictive proxies for our unobservable determinants $\ddot{z}$; and (4) variables that are not viably linked by economic reasoning to $\ddot{z}$ and which are therefore not legitimate proxies.

7 Testing Conditional Exogeneity

Although $\ddot{Z}_t$ is not observable, the economist is often able to observe variables that are fundamentally not determined by the natural experiment and that may serve as proxies for $\ddot{Z}_t$. These proxies can be used to empirically test CIPP.

We begin by assuming that we have a set of predictive proxies $X_t \equiv (\tilde{Z}_t, \tilde{W}_t)$ that can be used to construct $X_{\psi t}$ as previously defined. (Note that for present purposes ψ may be chosen differently than before, but we keep the notation the same for simplicity.) We explicitly allow the possibility that X_t may be null so that $X_{\psi t}$ contains only the constant.

Suppose that we also have some additional proxies for $\ddot{Z}_t$, say $\tilde{V}_t$, not already included in X_t. For example, in the cartel natural experiment, X_t may be null, and we may have some raw material prices and some demand shifters available as proxies $\tilde{V}_t$ for $\ddot{Z}_t$. Alternatively, we may have X_t already containing certain raw material prices and demand shifters as predictive proxies, but we may have further proxies for the same or different raw material prices and demand shifters not yet included in X_t.

In order to isolate any failure of conditional exogeneity, we require certain identifying information. We adopt the following structure as a means of plausibly achieving the required restrictions. Specifically, we suppose the additional proxies $\tilde{V}_t$ are related to $\ddot{Z}_t$ according to

$$\tilde{V}_t = v(\ddot{Z}_t, \ddot{U}_t), \qquad t = 1, 2, \ldots,$$

for some unknown function v and unobservable variables $\ddot{U}_t$, conveniently viewed as measurement errors. Thus, v specifies that the proxies are determined by $\ddot{Z}_t$ and some measurement errors $\ddot{U}_t$ in the same way for all observations.

The properties we require of $\ddot{U}_t$ are: (1) Λ_t is fundamentally non-determining for $\ddot{U}_t$; (2) $\ddot{U}_t$ is fundamentally non-determining for Y_t. (The latter property distinguishes elements of $\ddot{U}_t$ from those of $\ddot{Z}_t$.) These are reasonable conditions for measurement errors to satisfy. With the conditions stated so far, it follows that Λ_t is fundamentally non-determining for $\tilde{V}_t$, so that $\tilde{V}_t$ satisfies the requirements of B.1 for predictive proxies.

With one further condition we have a means of identifying and testing failures of conditional exogeneity. This condition is that

$$\ddot{U}_t \perp \Lambda_t \mid \ddot{Z}_t, \tilde{Z}_t, \tilde{W}_t,$$

that is, the conditional distribution of $\ddot{U}_t$ given $\ddot{Z}_t, \tilde{Z}_t$, and $\tilde{W}_t$ is the same for all t. Again, this is a plausible condition for measurement errors to satisfy. Letting $J(\cdot|\ddot{z}, \tilde{z}, \tilde{w})$ denote this common distribution for $\ddot{U}_t$ given $(\ddot{Z}_t, \tilde{Z}_t, \tilde{W}_t) = (\ddot{z}, \tilde{z}, \tilde{w})$, it follows that the conditional expectation of $\tilde{V}_t$ given X_t is given in regime 0 by

$$\int v(\ddot{z}, \ddot{u}) \, dJ(\ddot{u} \mid \ddot{z}, x) \, d\breve{G}_0(\ddot{z} \mid x)$$

and in regime 1 by

$$\int v(\ddot{z}, \ddot{u}) \, dJ(\ddot{u} \mid \ddot{z}, x) \, d\breve{G}_1(\ddot{z} \mid x).$$

If indeed conditional exogeneity ($\breve{G}_0 = \breve{G}_1$) holds, then the conditional expectation of $\tilde{V}_t$ given X_t is identical across regimes. That is, the stability of this conditional expectation is *necessary* for conditional exogeneity. Thus, if we find empirical evidence that this stability fails, then we have evidence of the failure of conditional exogeneity.

Observe the roles played by our assumptions on $v, \ddot{U}_t$, and J. If v depended on Λ_t, if $\ddot{U}_t$ were determined by Λ_t, or if J were not stable across regimes, we would also have the conditional expectation of $\tilde{V}_t$ given X_t differing across regimes and we could not isolate the failure of conditional exogeneity as the sole and necessary reason for this instability.

We emphasize that the stability of this conditional expectation is a necessary and not a sufficient condition for conditional exogeneity. Thus, tests based on this stability indicator could fail to detect violations of conditional exogeneity. It is nevertheless possible to gain additional power against failures of conditional exogeneity by testing further necessary conditions. In particular, let ϕ be any measurable $r \times 1$ vector-valued function taking $\tilde{V}_t$ as an argument. Then we also have that the conditional expectation of $\phi(\tilde{V}_t)$ given X_t is stable across regimes given conditional exogeneity; the common conditional expectation is

$$\int \phi(v(\ddot{z},\ddot{u}))\, dJ(\ddot{u}\,|\,\ddot{z},x)\, d\breve{G}(\ddot{z}\,|\,x).$$

For discussion of useful choices of ϕ see Stinchcombe and White (1998).

To construct a test of the stability of this conditional expectation, we again exploit the flexible approximation capabilities of $X_{\psi t}$. Thus, we can assume that there exist matrices $a^*_{\phi,0}$ and $a^*_{\phi,1}$ such that

$$\begin{aligned} \int \phi(v(\ddot{z},\ddot{u}))' dJ(\ddot{u}\,|\,\ddot{z},x)\, d\breve{G}_0(\ddot{z}|x) &= x'_\psi a^*_{\phi,0} \\ \int \phi(v(\ddot{z},\ddot{u}))' dJ(\ddot{u}\,|\,\ddot{z},x)\, d\breve{G}_1(\ddot{z}\,|\,x) &= x'_\psi a^*_{\phi,1}. \end{aligned}$$

Under the conditional exogeneity null hypothesis

$$H_0 : \breve{G}_0 = \breve{G}_1 \qquad a.s.$$

we have $a^*_{\phi,0} = a^*_{\phi,1}$, and it is this necessary condition that we test.

A straightforward analog of the Chow (1960) test statistic can be constructed using the regime 0 and regime 1 OLS estimators,

$$\begin{aligned} \hat{\alpha}_{\phi,0} &\equiv (\sum_{t=1}^{n}(1-\Lambda_t)X_{\psi t}X'_{\psi t})^{-1} \sum_{t=1}^{n}(1-\Lambda_t)X_{\psi t}\, \phi(\tilde{V}_t)' \\ \hat{\alpha}_{\phi,1} &\equiv (\sum_{t=1}^{n}\Lambda_t X_{\psi t}X_{\psi t})^{-1} \sum_{t=1}^{n}\Lambda_t X_{\psi t}\, \phi(\tilde{V}_t)'. \end{aligned}$$

Using $\hat{\alpha}_{\phi,0}$ and $\hat{\alpha}_{\phi,1}$ we can form a Wald (1943) version of the Chow test,

$$\mathcal{W}_n \equiv n\ [\text{vec}\ (\hat{\alpha}_{\phi,0} - \hat{\alpha}_{\phi,1})]'\hat{C}_n^{-1}\ [\text{vec}\ (\hat{\alpha}_{\phi,0} - \hat{\alpha}_{\phi,1})],$$

where $\hat{C}_n$ is a consistent estimator of the asymptotic covariance matrix of

$$\sqrt{n}\ \text{vec}\ (\hat{\alpha}_{\phi,0} - \hat{\alpha}_{\phi,1} - (a^*_{\phi,0} - a^*_{\phi,1})).$$

Under H_0, $\mathcal{W}_n$ has the $\mathcal{X}^2_{qr}$ distribution asymptotically.

Equivalently, a convenient test statistic can be computed by exploiting the representation

$$\phi(\tilde{V}_t)' = X'_{\psi t}\, a^*_{\psi,0} + \Lambda_t X'_{\phi t}(a^*_{\phi,1} - a^*_{\phi,0}) + \eta'_t \qquad t = 1, 2, \ldots,$$

where η_t is uncorrelated by construction with $X_{\psi t}$ and $\Lambda_t X_{\psi t}$. Using the vec operator and the identity vec $(ABC) = (C' \otimes I)$ vec B we have

$$\text{vec}\ \phi(\tilde{V}_t)' = (I \otimes X'_{\psi t})\ \text{vec}\ a^*_{\phi,0} + (I \otimes \Lambda_t X'_{\psi t})\ \text{vec}\ (a^*_{\phi,1} - a^*_{\phi,0}) +\ \text{vec}\ \eta'_t,$$

or

$$\phi(\tilde{V}_t) = \tilde{Z}'_{\psi t}\, \tilde{a}^*_{\phi,0} + \Lambda_t \tilde{Z}'_{\psi t}(\tilde{a}^*_{\phi,1} - \tilde{a}^*_{\phi,0}) + \eta_t, \qquad t = 1, 2, \ldots,$$

where

$$\begin{aligned}\tilde{Z}_{\psi t} &\equiv (I \otimes X_{\psi t}) \\ \tilde{a}^*_{\phi,0} &\equiv \text{vec } a^*_{\phi,0} \\ \tilde{a}^*_{\phi,1} &\equiv \text{vec } a^*_{\phi,1}.\end{aligned}$$

From this expression,we see that to test conditional exogeneity it suffices to regress $\phi(\tilde{V}_t)$ on $\tilde{Z}_{\psi t}$ and $\Lambda_t \tilde{Z}_{\psi t}$ and test whether the coefficients on $\Lambda_t \tilde{Z}_{\psi t}$ (that is, $\tilde{a}^*_{\phi,1} - \tilde{a}^*_{\phi,0}$) are jointly zero. This is quite straightforward, but care should be exercised to use a suitable heteroskedasticity-and-autocorrelation consistent (HAC) covariance estimator in constructing the test statistic.

We formalize our discussion by stating the following assumptions and results.

Assumption C.1 (Data Generating Process) *Assumption B.1 holds and additional observed data are generated from a realization of* $(\tilde{V}_t, \ddot{U}_t, \ddot{Z}_t)$, $t = 1, 2, \ldots$, *where* $(\ddot{U}_t, \ddot{Z}_t)$ *stably isolates* Λ_t *for* $\tilde{V}_t$ *according to*

$$\tilde{V}_t = v(\ddot{U}_t, \ddot{Z}_t)$$

for some unknown measurable vector-valued function v, *where* $\ddot{U}_t$ *is fundamentally non-determining for* Y_t.

For all $t = 1, 2, \ldots$, *the conditional distribution of* $\ddot{U}_t$ *given* $(\ddot{Z}_t, \tilde{Z}_t, \tilde{W}_t) = (\ddot{Z}_t, X_t) = (\ddot{z}, x)$ *is given by* $J(\cdot \,|\, \ddot{z}, x)$.

This assumption formalizes our earlier discussion concerning the generation of proxies $\tilde{V}_t$. Next we impose suitable moment conditions.

Assumption C.2 (Finiteness of Moments) *For given* $q \in \mathbb{N}$ *and given known measurable scalar-valued functions* $\psi_0 = 1$, ψ_j, $j = 1, \ldots, q-1$, *let* $x_\psi \equiv (\psi_0(x), ..., \psi_{q-1}(x))'$. *Let* ϕ *be a given known measurable* $r \times 1$ *vector-valued function. Assume*

(a) $0 < p_1 < 1$;

(b) (i) $M_{\psi,0} \equiv \int x_\psi x_\psi' \, d\breve{H}_0(x) < \infty, \quad \det M_{\psi,0} > 0$;

(ii) $M_{\psi,1} \equiv \int x_\psi x_\psi' \, d\breve{H}_1(x) < \infty, \quad \det M_{\psi,1} > 0$;

(c) (i) $K_{\psi\phi,0} \equiv \int x_\psi \phi(v(\ddot{u}, \ddot{z}))' \, dJ(\ddot{u} \,|\, \ddot{z}, x) \, d\breve{F}_0(\ddot{z}, x) < \infty$;

(ii) $K_{\psi\phi,1} \equiv \int x_\psi \phi(v(\ddot{u}, \ddot{z}))' \, dJ(\ddot{u} \,|\, \ddot{z}, x) \, d\breve{F}_1(\ddot{z}, x) < \infty$.

This assumption ensures the existence of the coefficient matrices

$$a^*_{\phi,0} \equiv M^{-1}_{\psi,0} K_{\psi\phi,0}, \qquad a^*_{\phi,1} \equiv M^{-1}_{\psi,1} K_{\psi\phi,1}.$$

Using these coefficient matrices, we can also define the residual vector

$$\eta_t' \equiv \phi(\tilde{V}_t)' - X'_{\psi t} a^*_{\phi,0} - \Lambda_t X'_{\psi t}(a^*_{\phi,1} - a^*_{\phi,0}), \qquad t = 1, 2, \ldots \,.$$

By construction, these residuals are uncorrelated with both $X_{\psi t}$ and $\Lambda_t X_{\psi t}$. The desired behavior of our test statistic is ensured by the following condition, in which we define

$$\hat{M}_{\psi,1} \equiv T_1^{-1} \sum_{t=1}^{n} \Lambda_t X_{\psi t} X'_{\psi t}.$$

Assumption C.3 (Laws of Large Numbers, Central Limit Theorem)

(a) $\hat{p}_1 \xrightarrow{p} p_{1;}$

(b) $\hat{M}_{\psi,0} \xrightarrow{p} M_{\psi,0}$ *and* $\hat{M}_{\psi,1} \xrightarrow{p} M_{\psi,1}$;

(c) *There exists a* $2qr \times 2qr$ *matrix* D, *finite and non-singular, such that*

$$D^{-1/2} n^{-1/2} \sum_{t=1}^{n} \left[\begin{array}{c} I \otimes X_{\psi t} \\ \Lambda_t (I \otimes X_{\psi t}) \end{array} \right] \eta_t \xrightarrow{d} N(0, I).$$

The conditional exogeneity test statistic is defined as

$$\mathcal{W}_n \equiv n[\text{ vec } (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0})]' \hat{C}_n^{-1} [\text{ vec } (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0})],$$

where

$$\begin{array}{rcl} \hat{C}_n & \equiv & R \hat{M}_{\psi}^{-1} \hat{D}_n \hat{M}_{\psi}^{-1} R', \\ R & = & [\mathbf{0}_{qr}, I_{qr}], \\ \hat{M}_{\psi} & = & \left[\begin{array}{cc} (1 - \hat{p}_1) \hat{M}_{\psi,0} + \hat{p}_1 \hat{M}_{\psi,1} & \hat{p}_1 \hat{M}_{\psi,1} \\ \hat{p}_1 \hat{M}_{\psi,1} & \hat{p}_1 \hat{M}_{\psi,1} \end{array} \right], \end{array}$$

and $\hat{D}_n$ is a consistent estimator of D. For example, in the absence of dynamic misspecification, a heteroskedasticity-consistent estimator for D is

$$\hat{D}_n = n^{-1} \sum_{t=1}^{n} \left[\begin{array}{c} I \otimes X_{\psi t} \\ \Lambda_t (I \otimes X_{\psi t}) \end{array} \right] \hat{\eta}_t \hat{\eta}'_t \left[\begin{array}{c} I \otimes X_{\psi t} \\ \Lambda_t (I \otimes X_{\psi t}) \end{array} \right]',$$

where

$$\hat{\eta}'_t \equiv \phi(\tilde{V}_t)' - X'_{\psi t} \, \hat{\alpha}_{\phi,0} - \Lambda_t X'_{\psi t} (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0}).$$

We have the following result, establishing the properties of our test for conditional exogeneity given predictive proxies.

Proposition 7.1: *Suppose Assumptions C.1-C.3 hold, and that* $\hat{D}_n \xrightarrow{p} D$. *Suppose further that for finite matrices* $a^*_{\phi,0}$ *and* $a^*_{\phi,1}$

$$\begin{array}{rcl} \int \phi(v(\ddot{z}, \ddot{u}))' dJ(\ddot{u}|\ddot{z}, x) \, d\breve{G}_0(\ddot{z}|x) & = & x'_{\psi} a^*_{\phi,0} \\ \int \phi(v(\ddot{z}, \ddot{u}))' dJ(\ddot{u}|\ddot{z}, x) \, d\breve{G}_1(\ddot{z}|x) & = & x'_{\psi} a^*_{\phi,1}. \end{array}$$

(i) *Given* $H_0 : \breve{G}_0 = \breve{G}_1$ *a.s., that is,* $\ddot{Z}_t \perp \Lambda_t \mid X_t$*, we have* $a^*_{\phi,0} = a^*_{\phi,1}$*, and as* $n \to \infty$

$$\mathcal{W}_n \xrightarrow{d} \mathcal{X}^2_{qr}.$$

(ii) *Suppose instead that* $a^*_{\phi,0} \neq a^*_{\phi,1}$*. Then for any sequence* $\{k_n\} = o(n)$

$$P[\mathcal{W}_n > k_n] \to 1 \quad as \quad n \to \infty.$$

Proposition 7.1(i) implies that we can perform a test of conditional exogeneity having asymptotic level α by rejecting H_0 whenever

$$\mathcal{W}_n \geq \mathcal{X}^2_{qr,1-\alpha},$$

where $\mathcal{X}^2_{qr,1-\alpha}$ is the $1-\alpha$ percentile of the $\mathcal{X}^2_{qr}$ distribution. Proposition 7.1(ii) implies that when conditional exogeneity fails in such a way that $a^*_{\phi,0} \neq a^*_{\phi,1}$, then a test of fixed asymptotic level α can detect this with probability approaching one. Further, the asymptotic level can be driven to zero, and, provided this is at a proper rate such that $\{k_n\} = o(n)$ (not too fast), power still approaches one.

These results are entirely standard. Local power properties are also entirely standard, so we do not state them here. They can be obtained by specializing results of White (1994, ch. 8).

To complete this section, we consider the consequences of rejecting or of failing to reject H_0. If one rejects H_0, then conditional exogeneity fails and the consistency of $\hat{\alpha}_\psi$ as an estimator for the average effect Δ^*_1 of the natural experiment is seriously in question. If one's objective is to undermine a rival theory, then this rejection accomplishes that goal by demonstrating the existence of important determining factors (proxied by $\tilde{V}_t$) left out of account whose distribution differs between regimes and that may therefore account for some or all of the observed differences in outcomes between regimes.

More constructively, faced with rejection of H_0, one may append some or all of the elements of $\tilde{V}_t$ to $\tilde{W}_t$, thereby augmenting the set of predictive proxies. (Recall that our assumptions ensure that the elements of $\tilde{V}_t$ constitute valid predictive proxies satisfying B.l.) One may repeat the conditional exogeneity test with the augmented set of predictive proxies, continuing to augment $\tilde{W}_t$ with elements of $\tilde{V}_t$ until H_0 is no longer rejected or all available proxies have been utilized. In considering which proxies to move from $\tilde{V}_t$ to $\tilde{W}_t$ first, it makes sense to pay primary attention to proxies for the most important unobserved factors and/or for the unobserved factors whose distributions apparently differ the most between regimes. Thus, in the cartel example the price of a raw material that plays a major role in the variable cost of the product of interest should be given high priority. This priority is enhanced if the distribution of this raw material appears to change dramatically between regimes, but it is diminished if its distribution is relatively stable.

Finally, consider the consequences of failing to reject conditional exogeneity. then one has empirical evidence that accords with the consistency of $\hat{\alpha}_\psi$ as an estimator of the average effect Δ_1^* of the natural experiment. By constructing the test in such a way to achieve non-trivial power (by compelling choice of proxies and relevant transformations ϕ), such empirical evidence can considerably buttress one's claim to have a useful measure of the effect of interest.

We note, although for brevity we do not pursue this here, that even when H_0 is not rejected it may be helpful to augment $\tilde{W}_t$ with some or all of the elements of $\tilde{V}_t$. The reason for this is that by including additional valid predictive proxies in the prediction equation, one reduces the variation of the prediction error, which leads to more precise estimates of the effect of interest.

8 Concluding Remarks

In this paper we have considered methods for analyzing the effects of a natural experiment, with particular attention to natural experiments created by an intervention or structural change occurring at a specific point in time. Our analysis draws on the extensive literature on treatment effects (Rubin, 1974; Rosenbaum and Rubin, 1983; Hahn, 1998; Hirano, Imbens, and Ridder, 2003), but unlike the typical situation in the treatment effects literature, the treatment constituted by the natural experiments considered here may precede the measurement of relevant covariates. Accordingly, we have devoted particular attention to articulating a framework, provided in Section 2, that permits the effects of such treatments to be analyzed without introducing confounding biases. Our framework also has utility in the more traditional cross-section analysis of treatment effects, in that by requiring the explicit identification of observable and unobservable determining factors for the response variable of interest, it provides an explicit and extensive role for economic theory in identifying suitable and unsuitable covariates.

Given the major role heretofore played by the dummy variable approach in estimating the effect of such natural experiments as government policy interventions, merger events, and cartels, we provide a detailed examination of the conditions necessary and sufficient for the dummy variable approach to yield reliable estimates of the effects of interest. We find that the conditions under which the dummy variable approach delivers consistent estimates are very stringent. The dummy variable model must include all relevant determining factors, measured without error, and must either be correctly specified in full or, if misspecified, must be applied in a setting in which the joint distribution of the observable and unobservable determining factors does not change between the regimes of the natural experiment. Neither of these possibilities is particularly plausible for most phenomena of interest to economists.

We then turn our attention to methods that draw on the extensive treatment effects literature. Using notions of unconfoundedness (Rubin, 1974) and the propensity score (Rosenbaum and Rubin, 1983), Hahn (1998) and Hirano, Imbens, and Ridder (2003) have proposed estimators for treatment effects in cross-section settings that attain the semiparametric efficiency bound. These estimators are consistent and asymptotically normal estimators for the effects of interest, and in particular for the "treatment effect on the treated," without having to impose correct specification of an underlying model, without requiring the absolutely accurate measurement of all determining factors, and without requiring the stability of the joint distribution of the determining factors between treatment regimes.

The Hahn and HIR estimators are thus potentially appealing candidates for estimating the effects of the natural experiments of interest here, but they also possess certain drawbacks. First, they are potentially computationally challenging, as they involve nonparametric estimation of one or more unknown conditional expectations. Second, their properties in the time-series applications of interest here are presently unknown. One may expect that plausible conditions can be found under which the Hahn and HIR estimators retain their asymptotic normality properties in time-series settings relevant in economic applications, but the semiparametric bound for these settings is currently unknown, as is whether the Hahn or HIR estimators achieve this bound.

It is of clear interest to extend the analysis of the Hahn and HIR estimators to the time-series context, but here we have pursued the less ambitious but still useful goal of proposing and analyzing a quasi-nonparametric estimator that shares many of the advantages of the Hahn and HIR estimators, but which is computationally straightforward and whose asymptotic properties can be straightforwardly analyzed under conditions that plausibly hold in either time-series or cross-section applications in economics. Our estimator can be computed from a dummy variable regression similar to the simple dummy variable approach first analyzed, but as for the Hahn and HIR estimators, the requirements of correct specification, absence of measurement error, etc., are no longer an issue. As we show, this estimator is consistent for the effect of interest and asymptotically normal under conditions that admit considerable time dependence, with an asymptotic variance that can be straightforwardly estimated by a variety of convenient HAC estimators. Because of its computational simplicity and known properties, this alternate quasi-nonparametric estimator should prove useful in applications.

As mentioned above, the framework of Section 2 provides clear opportunities for gaining insight into the choice of covariates. In Sections 6 and 7, we examine these issues in detail, from a theoretical perspective in Section 6 and from an empirical perspective in Section 7, where we provide tests for the conditional exogeneity property.

There are a host of interesting questions and opportunities for future research posed by the work reported here. The following is intended as only a partial list. Determining the properties of the Hahn and HIR estimators (as well as other natural variants) in a time-series setting is of immediate interest. Although the conditions for the DGP specified here permit considerable time-series dependence, they do not apply generally to cointegrated processes. Accordingly, it is of definite interest to extend the analysis undertaken here to permit cointegrated DGPs. For this, it will be necessary to pursue another interesting direction for further research, which is to go beyond the analysis of the determining reduced form that has been our focus here and to analyze the determining structural equations. This will permit a decomposition of the total effect of a natural experiment into component direct and indirect effects. In particular, this will permit researchers to work with time-series models that explicitly include lagged dependent variables or cointegrating effects. Another interesting possibility is to build on the framework developed here to permit estimation of the effects of continuous treatments, analogous to the analysis for binary treatment conducted here.

On the empirical side, the methods proposed here offer opportunities for gaining considerable new or additional insight into a variety of economic phenomena involving natural experiments operating in time, and we look forward to pursuing these in future work.

9 Mathematical Appendix

Proof of Proposition 3.1. The result follows immediately by applying, e.g., Proposition 2.27 of White (2001) to $\hat{\alpha}$ and $\hat{\beta}$.

Proof of Proposition 3.2 By definition

$$\begin{aligned}\alpha^* - \Delta_1^* &= \mu_1 - m_1'\beta^* - (\mu_1 - \mu_{01}) \\ &= \mu_{01} - m_1'\beta^* \\ &= \mu_{01} - m_1'\beta_0^* + m_1'(\beta_0^* - \beta^*).\end{aligned}$$

Now

$$\mu_{01} \equiv \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}).$$

It follows straightforwardly that

$$\mu_{01} = m_1'\beta_{01}^*,$$

as β_{01}^* satisfies the orthogonality conditions

$$\int z(c_0(\tilde{z}, \ddot{z}) - z'\beta_{01}^*)\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) = 0,$$

and because the first element of z is unity, this implies

$$\int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) - \int z'\beta_{01}^*\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) = \mu_{01} - m_1'\beta_{01}^* = 0.$$

Substituting $m_1'\beta_{01}^*$ for μ_{01} gives

$$\alpha^* - \Delta_1^* = m_1'(\beta_{01}^* - \beta_0^*) + m_1'(\beta_0^* - \beta^*).$$

To obtain the first expression of Proposition 3.2, it suffices to show that

$$\begin{aligned}m_1'\beta^* &= m_1'\beta_0^* - p_1(m_0 - m_1)'S'S(\beta_1^* - \beta_0^*) \\ &\quad -p_1(1-p_1)(m_0 - m_1)'S'\tilde{M}^{-1}(\tilde{M}_1 - \tilde{M}_0)S(\beta_1^* - \beta_0^*).\end{aligned}$$

We begin by decomposing

$$
\begin{aligned}
\beta^* &\equiv [(1-p_1)M_0 + p_1(M_1 - m_1 m_1')]^{-1}[(1-p_1)L_0 + p_1(L_1 - m_1\mu_1)] \\
&= [M_0 + \pi(M_1 - m_1 m_1')]^{-1}[L_0 + \pi(L_1 - m_1\mu_1)],
\end{aligned}
$$

where $\pi \equiv p_1/(1-p_1)$. Partitioning M_0 and M_1 to explicitly expose the non-constant components of $z = (1, \tilde{z})'$, we have

$$
\begin{aligned}
M_0 &= \begin{bmatrix} 1 & m_0' S' \\ S m_0 & S M_0 S' \end{bmatrix} \\
M_1 &= \begin{bmatrix} 1 & m_1' S' \\ S m_1 & S M_1 S' \end{bmatrix},
\end{aligned}
$$

where S is the selection matrix such that $\tilde{z}' = Sz$. It follows that

$$
M_0 + \pi(M_1 - m_1 m_1') = \begin{bmatrix} 1 & m_0' S' \\ S m_0 & S(M_0 + \pi(M_1 - m_1 m_1'))S' \end{bmatrix},
$$

as the first row and column of $M_1 - m_1 m_1'$ are zero.

Applying the formula for the partitioned inverse gives

$$
[M_0 + \pi(M_1 - m_1 m_1')]^{-1} = \begin{bmatrix} 1 + m_0' S' \ddot{M}^{-1} S m_0 & -m_0' S' \ddot{M}^{-1} \\ -\ddot{M}^{-1} S m_0 & \ddot{M}^{-1} \end{bmatrix},
$$

where

$$
\begin{aligned}
\ddot{M} &\equiv \tilde{M}_0 + \pi \tilde{M}_1 \\
\tilde{M}_0 &\equiv S(M_0 - m_0 m_0')S' \\
\tilde{M}_1 &\equiv S(M_1 - m_1 m_1')S'.
\end{aligned}
$$

Next we have

$$
L_0 + \pi(L_1 - m_1\mu_1) = \begin{bmatrix} \mu_0 \\ S[L_0 + \pi(L_1 - m_1\mu_1) \end{bmatrix},
$$

so that

$$
\beta^* = \begin{bmatrix} \mu_0 + m_0' S' \ddot{M}^{-1} S m_0 \mu_0 - m_0' S' \ddot{M}^{-1} S[L_0 + \pi(L_1 - m_1\mu_1)] \\ -\ddot{M}^{-1} S m_0 \mu_0 + \ddot{M}^{-1} S[L_0 + \pi(L_1 - m_1\mu_1) \end{bmatrix}.
$$

Because the first element of m_1 is unity, from this we obtain

$$
\begin{aligned}
m_1'\beta^* = {} & \mu_0 + m_0' S' \ddot{M}^{-1} S m_0\, \mu_0 - m_0' S' \ddot{M}^{-1} S[L_0 + \pi(L_1 - m_1\mu_1)] \\
& - m_1' S' \ddot{M}^{-1} S m_0 \mu_0 + m_1' S' \ddot{M}^{-1} S[L_0 + \pi(L_1 - m_1\mu_1)].
\end{aligned}
$$

The definitions of β_0^* and β_1^* imply that

$$
\begin{aligned}
\tilde{M}_0 S \beta_0^* &= S(L_0 - m_0\mu_0) \\
\tilde{M}_1 S \beta_1^* &= S(L_1 - m_1\mu_1).
\end{aligned}
$$

Substituting these expressions gives

$$\begin{aligned}
m_1'\beta^* &= \mu_0 - (m_0 - m_1)'S'\ddot{M}^{-1}(\tilde{M}_0 S\beta_0^* + \pi\tilde{M}_1 S\beta_1^*) \\
&= \mu_0 - (m_0 - m_1)'S'\tilde{M}^{-1}[(1-p_1)\tilde{M}_0 S\beta_0^* + p_1\tilde{M}_1 S\beta_1^*],
\end{aligned}$$

where we replace $\ddot{M}$ with $\ddot{M} = (1-p_1)^{-1}\tilde{M}$. Adding and subtracting terms appropriately gives

$$\begin{aligned}
m_1'\beta^* &= \mu_0 - (1-p_1)(m_0 - m_1)'S'S\beta_0^* - p_1(m_0 - m_1)'S'S\beta_1^* \\
&\quad -(m_0 - m_1)'S\tilde{M}^{-1}[(1-p_1)(\tilde{M}_0 - \tilde{M})S\beta_0^* + p_1(\tilde{M}_1 - \tilde{M})S\beta_1^*] \\
&= \mu_0 - (m_0 - m_1)'S'S\beta_0^* - p_1(m_0 - m_1)'S'S(\beta_1^* - \beta_0^*) \\
&\quad -(m_0 - m_1)'S'\tilde{M}^{-1}[(1-p_1)p_1(\tilde{M}_0 - \tilde{M}_1)S\beta_0^* + p_1(1-p_1)(\tilde{M}_1 - \tilde{M}_0)S\beta_1^*],
\end{aligned}$$

after some rearrangement and using the facts that

$$\begin{aligned}
\tilde{M}_0 - \tilde{M} &= p_1(\tilde{M}_0 - \tilde{M}_1) \\
\tilde{M}_1 - \tilde{M} &= (1-p_1)(\tilde{M}_1 - \tilde{M}_0).
\end{aligned}$$

The orthogonality conditions underlying β_0^* ensure that $\mu_0 = m_0'\beta_0^*$. It follows that

$$\begin{aligned}
\mu_0 - (m_0 - m_1)'S'S\beta_0^* &= m_0'\beta_0^* - (m_0 - m_1)'S'S\beta_0^* \\
&= m_1'\beta_0^* + (m_0 - m_1)'\beta_0^* - (m_0 - m_1)'S'S\beta_0^* \\
&= m_1'\beta_0^*,
\end{aligned}$$

The last equality holds as the first element of $m_0 - m_1$ is zero. Thus

$$\begin{aligned}
m_1'\beta^* &= m_1'\beta_0^* - p_1(m_0 - m_1)'S'S(\beta_1^* - \beta_0^*) \\
&\quad -p_1(1-p_1)(m_0 - m_1)'S'\tilde{M}^{-1}(\tilde{M}_1 - \tilde{M}_0)S(\beta_1^* - \beta_0^*),
\end{aligned}$$

as was to be shown, establishing the first expression of Proposition 3.2.

To obtain the second expression, it suffices to show that

$$\begin{aligned}
m_1'(\beta_{01}^* - \beta_0^*) &= \int c_0(\tilde{z}, \ddot{z})(d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z}))\, dH_1(\tilde{z}) \\
&\quad + \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})(dH_1(\tilde{z}) - dH_0(\tilde{z})) \\
&\quad +(m_0 - m_1)'\beta_0^*,
\end{aligned}$$

as $(m_0 - m_1)'S'S(\beta_1^* - \beta_0^*) = (m_0 - m_1)'(\beta_1^* - \beta_0^*)$. Simplifying, we have

$$\int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z} \mid \tilde{z})\, dH_1(\tilde{z}) - \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z}) + m_0'\beta_0^* - m_1'\beta_0^*$$

$$\begin{aligned}
&= \mu_{01} - m_0'\beta_0^* + m_0'\beta_0^* - m_1'\beta_0^* \\
&= m_1'\beta_{01}^* - m_1'\beta_0^*,
\end{aligned}$$

using the definition of μ_{01} and the orthogonality condition ensuring that $\mu_0 = m_0'\beta_0^*$. The desired result therefore holds, establishing the second expression for $\alpha^* - \Delta_1^*$ in Proposition 3.2.

Proof of Corollary 3.3. Immediate from of Proposition 3.2.

Proof of Corollary 3.4. (i) We have

$$\begin{aligned}
\beta_0^* &\equiv M_0^{-1} L_0 \\
&= M_0^{-1} \int z c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z}) \\
&= M_0^{-1} \int z(z'b^* + u_0(\ddot{z}))\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z}) \\
&= M_0^{-1}[\int zz' d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z})]b^* \\
&\quad + M_0^{-1} \int z u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, dH_0(\tilde{z}) \\
&= M_0^{-1} M_0 b^* + M_0^{-1} \int z[\int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})]\, dH_0(\tilde{z}) \\
&= b^*,
\end{aligned}$$

as we have assumed that $\int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z}) = 0$. Next,

$$\begin{aligned}
\beta_{01}^* &\equiv M_1^{-1} L_{01} \\
&= M_1^{-1} \int z c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= M_1^{-1} \int z(z'b^* + u_0(\ddot{z}))\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= M_1^{-1}[\int zz' d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z})]b^* \\
&\quad + M_1^{-1} \int z u_0(\ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= b^* + M_1^{-1} \int z u_0(\ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= b^* + M_1^{-1} \int z[\int u_0(\ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z}) - \int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})]\, dH_1(\tilde{z}) \\
&= b^* + M_1^{-1} \int z u_0(\ddot{z})[d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z})]\, dH_1(\tilde{z}),
\end{aligned}$$

where the next to last equality follows given $\int u_0(\ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z}) = 0$.

Finally,

$$\begin{aligned}
\beta_1^* &\equiv M_1^{-1}L_1 \\
&= M_1^{-1}L_{01} + M_1^{-1}(L_1 - L_{01}) \\
&= \beta_{01}^* + M_1^{-1}\int z(c_1(\ddot{z},\tilde{z}) - c_0(\ddot{z},\tilde{z}))\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= \beta_{01}^* + M_1^{-1}\int z\delta^* d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= \beta_{01}^* + M_1^{-1}\int zz' d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \;\; S_1'\delta^* \\
&= \beta_{01}^* + S_1'\delta^*.
\end{aligned}$$

(ii) If $\tilde{G}_0 = \tilde{G}_1$ a.s.$-H_1$, then

$$M_1^{-1}\int zu_0(\ddot{z})[d\tilde{G}_1(\ddot{z}|\tilde{z}) - d\tilde{G}_0(\ddot{z}|\tilde{z})]\; dH_1(\tilde{z}) = 0,$$

so $\beta_{01}^* = b^* = \beta_0^*$, $\beta_1^* - \beta_0^* = S_1'\delta^*$, and finally $S(\beta_1^* - \beta_0^*) = SS_1'\delta^* = 0$, as $SS_1' = 0$, so that from Proposition 3.2

$$\alpha^* = \Delta_1^*.$$

Moreover,

$$\begin{aligned}
\Delta_1^* &= \mu_1 - \mu_{01} \\
&= \int (c_1(\ddot{z},\tilde{z}) - c_0(\ddot{z},\tilde{z}))\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= \delta^* \int d\tilde{G}_1(\ddot{z}|\tilde{z})\, dH_1(\tilde{z}) \\
&= \delta^*,
\end{aligned}$$

so $\alpha^* = \delta^* = \Delta_1^*$.

Proof of Corollary 3.5.

(i) The results for β_0^* and β_{01}^* follow by reasoning identical to 3.4(i), replacing b^* with b_0^*. The result for β_1^* follows by reasoning analogous to that for β_0^* replacing subscript 0's with subscript 1's.

(ii) If $\tilde{G}_0 = \tilde{G}_1$ a.s.$-H_1$, then $\beta_0^* = \beta_{01}^*$ by the same logic as in 3.4(ii). The expression for $\alpha^* - \Delta_1^*$ now follows from Proposition 3.2, setting $\beta_0^* = \beta_{01}^*$, $\beta_0^* = b_0^*$, and $\beta_1^* = b_1^*$.

Proof of Proposition 4.1: Lemma 4.1 of Dawid (1979) ensures that $\ddot{Z} \perp \Lambda \mid (\tilde{Z},\tilde{W})$ implies $(\ddot{Z},\tilde{Z},\tilde{W}) \perp (\Lambda,\tilde{Z},\tilde{W}) \mid (\tilde{Z},\tilde{W})$. It follows from Dawid

(1979, Lemma 4.2) that $f(\ddot{Z}, \tilde{Z}, \tilde{W}) \perp g(\Lambda, \tilde{Z}, \tilde{W}) \mid (\tilde{Z}, \tilde{W})$ for any measurable functions f and g. Put $f(\ddot{z}, \tilde{z}, \tilde{w}) = (c_0(\tilde{z}, \ddot{z}),\ c_1(\tilde{z}, \ddot{z}))$ and $g(\lambda, \tilde{z}, \tilde{w}) = \lambda$. It follows that

$$(c_0(\tilde{Z}, \ddot{Z}), c_1(\tilde{Z}, \ddot{Z})) \perp \Lambda \mid (\tilde{Z}, \tilde{W}).$$

Proof of Proposition 5.1. The result follows immediately by applying, e.g., Proposition 2.27 of White (2001) to $\hat{\alpha}_\psi$ and $\hat{\beta}_\psi$.

Proof of Proposition 5.2. We have (with $x \equiv (\tilde{z}, \tilde{w})$)

$$\begin{aligned}
\alpha_\psi^* - \Delta_1^* &= \mu_1 - m'_{\psi,1}\beta^*_{\psi,0} - (\mu_1 - \mu_{01}) \\
&= \mu_{01} - m'_{\psi,1}\beta^*_{\psi,0} \\
&= \int \breve{\mu}_{01}(x)\, d\breve{H}_1(x) - \int x'_\psi \beta^*_{\psi,0}\, d\breve{H}_1(x) \\
&= \int (\breve{\mu}_{01}(x) - \breve{\mu}_0(x))\, d\breve{H}_1(x) \\
&\quad - \int (\breve{\mu}_0(x) - x'_\psi \beta^*_{\psi,0})\, d\breve{H}_1(x),
\end{aligned}$$

where the last equality follows by adding and subtracting $\int \breve{\mu}_0(x)\, d\breve{H}_1(x)$ and the next to last equality follows because

$$\mu_{01} = \int [\int c_0(x)\, d\breve{G}_1(\ddot{z}|x)]\, d\breve{H}_1(x) = \int \breve{\mu}_{01}(x)\, d\breve{H}_1(x).$$

Proof of Corollary 5.3. (i) We have

$$\begin{aligned}
\beta^*_{\psi,0} &\equiv M^{-1}_{\psi,0} L_{\psi,0} \\
&= M^{-1}_{\psi,0} \int x_\psi c_0(\tilde{z}, \ddot{z})\, d\breve{G}_0(\ddot{z}|x)\, d\breve{H}_0(x) \\
&= M^{-1}_{\psi,0} \int x_\psi \breve{\mu}_0(x)\, d\breve{H}_0(x) \\
&= M^{-1}_{\psi,0} \int x_\psi x'_\psi b^*_{\psi,0}\, d\breve{H}_0(x),
\end{aligned}$$

where the last equality follows given $\breve{\mu}_0(x) = x'_\psi b^*_{\psi,0}$ and the next to last equality follows by definition of $\breve{\mu}_0$. Thus

$$\beta^*_{\psi,0} = M^{-1}_{\psi,0} M_{\psi,0}\, b^*_{\psi,0} = b^*_{\psi,0}.$$

(ii) From (i) above and Proposition 4.2 it follows that

$$\alpha^*_\psi - \Delta^*_1 = \int (\breve{\mu}_{01}(x) - \breve{\mu}_0(x))\, d\breve{H}_1(x).$$

Now

$$\begin{aligned}\breve{\mu}_{01}(x) - \breve{\mu}_0(x) &= \int c_0(\tilde{z}, \ddot{z})(d\breve{G}_1(\ddot{z}|x) - d\breve{G}_0(\ddot{z}|x)) \\ &= 0 \quad a.s. \quad -\breve{H}_1,\end{aligned}$$

given that $\breve{G}_0 = \breve{G}_1$ a.s. $-\breve{H}_1$. Thus $\alpha_\psi^* = \Delta_1^*$.

Proof of Theorem 5.4: By definition

$$\begin{aligned}\sqrt{n}(\hat{\alpha}_\psi - \alpha_\psi^*) &= \sqrt{n}(\hat{\mu}_1 - \mu_1 - \hat{m}_{\psi,1}'\hat{\beta}_{\psi,0} + m_{\psi,1}'\beta_{\psi,0}^*) \\ &= \sqrt{n}(\hat{\mu}_1 - \mu_1) - \beta_{\psi,0}^{*\prime}\sqrt{n}(\hat{m}_{\psi,1} - m_{\psi,1}) \\ &\quad -m_{\psi,1}'\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*) \\ &\quad -(\hat{m}_{\psi,1} - m_{\psi,1})'\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*).\end{aligned}$$

Assumption B.3(b) ensures that $\hat{m}_{\psi,1} - m_{\psi,1} = o_p(1)$ and Assumptions B.1, B.2(a), B.2(b.i), B.2(c.i), B.3, and B.5 imply $\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*) = O_p(1)$, so that

$$\begin{aligned}\sqrt{n}(\hat{\alpha}_{\psi n} - \alpha_\psi^*) &= \sqrt{n}(\hat{\mu}_1 - \mu_1 - \beta_{\psi,0}^{*\prime}(\hat{m}_{\psi,1} - m_{\psi,1})) \\ &\quad -m_{\psi,1}'\sqrt{n}(\hat{\beta}_{\psi,0} - \beta_{\psi,0}^*) + o_p(1).\end{aligned}$$

Substituting for $\hat{\mu}_1$ and $\hat{m}_{\psi,1}$, we have

$$\begin{aligned}&\hat{\mu}_1 - \mu_1 - \beta_{\psi,0}^{*\prime}(\hat{m}_{\psi,1} - m_{\psi,1}) \\ &\quad = T_1^{-1}\sum_{t\in\mathcal{T}_1}(Y_t - X_{\psi t}'\beta_{\psi,0}^* - \mu_1 + m_{\psi,1}'\beta_{\psi,0}^*) \\ &\quad = T_1^{-1}\sum_{t\in\mathcal{T}_1}[Y_t - X_{\psi t}'\beta_{\psi,1}^* + X_{\psi t}'\beta_{\psi,1}^* - X_{\psi t}'\beta_{\psi,0}^* - \alpha_\psi^*] \\ &\quad = T_1^{-1}\sum_{t\in\mathcal{T}_1}[\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)],\end{aligned}$$

where $\varepsilon_{t1} \equiv Y_t - X_{\psi t}'\beta_{\psi,1}^*$, $\mu_1^*(X_{\psi t}) \equiv X_{\psi t}'\beta_{\psi,1}^*$, $\mu_0^*(X_{\psi t}) = X_{\psi t}'\beta_{\psi,0}^*$, and we have substituted $\alpha_\psi^* \equiv \mu_1 - m_{\psi,1}'\beta_{\psi,0}^*$ in the second equality. Observe that by construction ε_{t1} is uncorrelated with $X_{\psi t}$ and thus with $(\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)$ for $t \in \mathcal{T}_1$. It follows that

$$\begin{aligned}&\sqrt{n}(\hat{\mu}_1 - \mu_1 - \beta_{\psi,0}^{*\prime}(\hat{m}_{\psi,1} - m_{\psi,1})) \\ &\quad = (T_1/n)^{-1}n^{-1/2}\sum_{t=1}^{n}\Lambda_t[\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)] \\ &\quad = p_1^{-1}\ n^{-1/2}\sum_{t=1}^{n}\Lambda_t[\varepsilon_{t1} + (\mu_1^*(X_{\psi t}) - \mu_0^*(X_{\psi t}) - \alpha_\psi^*)] + o_p(1),\end{aligned}$$

where the last equality follows given Assumptions B.3(a) and B.5(c).

It follows from Assumptions B.1, B.2(a), B.2(b.i), B.2(c.i), B.3 and B.5(a) that

$$\begin{aligned}\sqrt{n}(\hat{\beta}_{\psi,0}-\beta^*_{\psi,0}) &= (T_0/n)^{-1}M^{-1}_{\psi,0}n^{-1/2}\sum_{t=1}^{n}(1-\Lambda_t)X_{\psi t}\varepsilon_{t0}+o_p(1)\\ &= p_0^{-1}M^{-1}_{\psi,0}n^{-1/2}\sum_{t=1}^{n}(1-\Lambda_t)X_{\psi t}\varepsilon_{t0}+o_p(1).\end{aligned}$$

Combining these results, we have

$$\sqrt{n}(\hat{\alpha}_\psi-\alpha^*_\psi)=n^{-1/2}\sum_{t=1}^{n}\xi_t+o_p(1),$$

where

$$\xi_t\equiv p_1^{-1}\Lambda_t[\varepsilon_{t1}+(\mu_1^*(X_{\psi t})-\mu_0^*(X_{\psi t})-\alpha^*_\psi)]-p_0^{-1}m'_{\psi,1}M^{-1}_{\psi,0}(1-\Lambda_t)X_{\psi t}\varepsilon_{t0.}$$

It now follows immediately from Assumption B.5(i) that

$$\sqrt{n}(\hat{\alpha}_\psi-\alpha^*_\psi)\xrightarrow{d}N(0,\sigma^2_\xi).$$

Proof of Proposition 6.1. First suppose that $\mathcal{Y}=g(\mathcal{Z})$ with probability 1 (w.p.1). Let $A\in\mathcal{B}(\mathbb{R}^k)$ and $B\in\mathcal{B}(\mathbb{R}^m)$, where $\mathcal{B}(\mathbb{R}^k)$ and $\mathcal{B}(\mathbb{R}^m)$ are the Borel σ-fields generated by the open sets of $\mathbb{R}^k$ and $\mathbb{R}^m$ respectively. Then

$$\begin{aligned}P[\mathcal{V}\in A,\mathcal{Y}\in B|\mathcal{Z}] &= P[\mathcal{V}\in A,g(\mathcal{Z})\in B|\mathcal{Z}]\\ &= P[\mathcal{V}\in A|\mathcal{Z}]\,1[g(\mathcal{Z})\in B]\\ &= P[\mathcal{V}\in A|\mathcal{Z}]\,P[g(\mathcal{Z})\in B|\mathcal{Z}]\\ &= P[\mathcal{V}\in A|\mathcal{Z}]\,P[\mathcal{Y}\in B|\mathcal{Z}],\end{aligned}$$

as $\mathcal{Y}=g(\mathcal{Z})$ with probability 1 if and only if for all $B\in\mathcal{B}(\mathbb{R}^m)$ we have $P[g(\mathcal{Z})\in B|\mathcal{Z}]=1[g(\mathcal{Z})\in B]$. As the result holds for arbitrary A and B, we have that $\mathcal{Y}\perp\mathcal{V}\,|\,\mathcal{Z}$.

Next suppose that $\mathcal{Y}\perp\mathcal{V}\,|\,\mathcal{Z}$ holds, but that $\mathcal{Y}=g(\mathcal{Z})$ w.p.1 is false. That is, for all g measurable $-\mathcal{B}(\mathbb{R}^\ell)/\mathcal{B}(\mathbb{R}^m)$

$$P[f(\mathcal{V},\mathcal{Z})=g(\mathcal{Z})]<1.$$

Pick any g measurable–$\mathcal{B}(\mathbb{R}^\ell)/\mathcal{B}(\mathbb{R}^m)$. Then $P[f(\mathcal{V},\mathcal{Z})=g(\mathcal{Z})]<1$ implies the existence of sets $A_g\in\mathcal{B}(\mathbb{R}^k)$, $B_g\in\mathcal{B}(\mathbb{R}^\ell)$ and $C_g\in\mathcal{B}(\mathbb{R}^m)$ such that $P[A_g\times B_g]>0$ and for all $(v,z)\in A_g\times B_g$ we have

$$1[f(v,z)\in C_g]\neq 1[g(z)\in C_g].$$

Thus for all $z \in B_g$, $P[B_g] > 0$,

$$\begin{aligned} P[\,f(\mathcal{V}, z) \in C_g | \mathcal{Z} = z] &\neq P[g(\mathcal{Z}) \in C_g | \mathcal{Z} = z] \\ &= 1[g(z) \in C_g], \end{aligned}$$

so that for all $z \in B_g$,

$$0 < P[f(\mathcal{V}, z) \in C_g \,|\, \mathcal{Z} = z] < 1.$$

Now, let $A_{g,z} = f^{-1}(C_g, z)$ be the pre-image of C_g under $f(\cdot, z)$, so for all z

$$\{\omega \in \Omega \,|\, f(\mathcal{V}(\omega), z) \in C_g\} = \{\omega \in \Omega \,|\, \mathcal{V}(\omega) \in A_{g,z}\}.$$

It follows that

$$\{\omega \,|\, f(W(\omega), z) \in C_g, \mathcal{V}(\omega) \in A_{g,z}\} = \{\omega \,|\, \mathcal{V}(\omega) \in A_{g,z}\},$$

so that

$$P[\mathcal{V} \in A_{g,z}, f(\mathcal{V}, z) \in C_g \,|\, \mathcal{Z} = z] = P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z].$$

By conditional independence we also have for all z

$$\begin{aligned} &P[\mathcal{V} \in A_{g,z}, f(\mathcal{V}, x) \in C_g \,|\, \mathcal{Z} = z] \\ &\quad = P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z]\, P[f(\mathcal{V}, z) \in C_g \,|\, \mathcal{Z} = z] \\ &\quad = P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z]^2, \end{aligned}$$

as $P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z] = P[f(\mathcal{V}, z) \in C_g \,|\, \mathcal{Z} = z]$.

Combining our last two results we have that for all z

$$P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z] = P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z]^2.$$

But for all $z \in B_g$, $P[B_g] > 0$, we have

$$0 < P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z] = P[f(\mathcal{V}, z) \in C_g \,|\, \mathcal{Z} = z] < 1,$$

which implies

$$P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z] \neq P[\mathcal{V} \in A_{g,z} \,|\, \mathcal{Z} = z]^2$$

for all $z \in B_g$. We thus have a contradiction, and this contradiction is obtained regardless of our choice for g. It follows that we cannot simultaneously have $\mathcal{Y} \perp \mathcal{V} \,|\, \mathcal{Z}$ and that $\mathcal{Y} = g(\mathcal{Z})$ w.p.1 is false. The proof is complete.

Proof of Proposition 7.1. (i) If $\breve{G}_0 = \breve{G}_1 \quad a.s. - \breve{H}_1$ it follows immediately that $a^*_{\phi,0} = a^*_{\phi,1}$. The result now follows immediately by application of Theorem 4.31 of White (2001). (ii) The conditions given ensure the consistency of $\hat{\alpha}_{\phi,0}$ for $a^*_{\phi,0}$ and of $\hat{\alpha}_{\phi,1}$ for $a^*_{\phi,1}$. The result follows by application of Theorem 8.16 of White (1994).

References

Angrist, J., 2002a, Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment, American Economic Review, 92, 1535-1558.

Angrist, J., 2002b, How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America's Second Generation, The Quarterly Journal of Economics, 117, 997-1038.

Auten, G. and R. Carroll, 1999, The Effect of Income Taxes on Household Income, The Review of Economics & Statistics, 81, 681-693.

Berry, S. and J. Waldfogel, 2001, Do Mergers Increase Product Variety? Evidence From Radio Broadcasting, The Quarterly Journal of Economics, 116, 1009-1025.

Bronars, S. and J. Grogger, 1994, The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment, American Economic Review, 84, 1141-1156.

Campos, J., D. Hendry, and H.-M. Krolzig, 2003, Consistent Model Selection by an Automatic Gets Approach, Oxford University Nuffield College Working Paper.

Chen, X., H. Hong, and A. Tarozzi, 2004, Semiparametric Efficiency in GMM Models of Nonclassical Measurement Error, Missing Data and Treatment Effects, Working Paper, New York University, Department of Economics.

Chen, X. and H. White, 1999, Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators, IEEE Transactions on Information Theory, 45, 628-691.

Chow, G., 1960, Tests of Equality Between Sets of Coefficients in Two Linear Regressions, Econometrica, 28, 591-605.

Cook, D., 2002, World War II and Convergence, The Review of Economics & Statistics, 84, 131-138.

Dawid, A.P., 1979, Conditional Independence in Statistical Theory, Journal of the Royal Statistical Society, Series B, 41, 1-31.

Dawid, A.P., 2000, Causal Inference Without Counterfactuals, Journal of the American Statistical Association, 95, 407-448.

Deacon, R. and J. Sonstelie, 1985, Rationing by Waiting and the Value of Time: Results from a Natural Experiment, Journal of Political Economy, 93, 627-647.

Deltas, G. and G. Kosmopoulou, 2004, 'Catalogue' vs 'Order-of-Sale' Effects in Sequential Auctions: Theory and Evidence from a Rare Book Sale, Economic Journal, 114, 28-54.

Fisher, F., 1980, Multiple Regression in Legal Proceedings, Columbia Law Review, 85, 702-736.

Florens J.P. and D. Fougère, 1996, Noncausality in Continuous Time, Econometrica, 65, 1195-1212.

Florens, J.P. and M. Mouchart, 1982, A Note on Non-causality, Econometrica, 50, 583-591.

Frech III, H.E., 1976, The Property Rights Theory of the Firm: Empirical Results from a Natural Experiment, Journal of Political Economy, 84, 143-52.

Froehlich, M., 2004, A Note On the Role of the Propensity Score for Estimating Average Treatment Effects, Econometric Reviews, 23, 167-174.

Gallant, A.R. and H. White, 1992, On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks, Neural Networks, 5, 129-138.

Gencay, R., F. Selchuk, and B. Whitcher, 2001, An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. New York: Academic Press.

Gonçalves, S. and H. White, 2005, Bootstrap Standard Error Estimates for Linear Regression, Journal of the American Statistical Association (in press).

Granger, C.W.J., 1969, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica, 37, 424-438.

Greene, W., 1993, Econometric Analysis. New York: MacMillan.

Hahn, J., 1998, On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects, Econometrica, 66, 315-331.

Heckman, J., H. Ichimura, and P. Todd, 1998, Matching as an Econometric Evaluation Estimator, Review of Economic Studies, 65, 261-294.

Hendry, D.F. and H.-M. Krolzig, 2001, Automatic Econometric Model Selection with PcGets. London: Timberlake Consultants Press.

Higgins, R. and P. Johnson, 2003, The Mean Effect of Structural Change on the Dependent Variable is Accurately Measured by the Intercept Change Alone, Economics Letters, 80, 255-259.

Hirano, K., G. Imbens, and G. Ridder, 2003, Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score, Econometrica, 71, 1161-1189.

Hornik, K., M. Stinchcombe, and H. White, 1989, Multilayer Feedforward Networks are Universal Approximators, Neural Networks, 2, 359-366.

Hornik, K., M. Stinchcombe, and H. White, 1990, Universal Approximation of an Unknown Mapping and Its Derivatives Using Multilayer Feedforward Networks, Neural Networks, 3, 551-560.

Hotz, V.J., C. Mullin, and S. Sanders, 1997, Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analysing the Effects of Teenage Childbearing, Review of Economic Studies, 64, 575-603.

Kandel, E. and B.Z. Zilberfarb, 1999, Differential Interpretation of Information in Inflation Forecasts, The Review of Economics & Statistics, 81, 217-226.

Metrick, A., 1995, A Natural Experiment in 'Jeopardy!', American Economic Review, 85, 240-253.

Meyer, B.D., 1995, Natural and Quasi-Experiments in Economics, Journal of Business & Economic Statistics, 13, 151-162.

Meyer, B.D., W.K. Viscusi, and D.L. Durbin, 1995, Workers' Compensation and Injury Duration: Evidence from a Natural Experiment, American Economic Review, 85, 322-340.

Millimet, D. and List, J., 2003, A Natural Experiment on the 'Race to the Bottom' Hypothesis: Testing for Stochastic Dominance in Temporal Pollution

Trends, Oxford Bulletin of Economics and Statistics, 65, 395-420.

Perez-Amaral, T., G.M. Gallo, and H. White, 2003, A Flexible Tool for Model Building: the Relevant Transformation of the Inputs Network Approach (RETINA), Oxford Bulletin of Economics & Statistics, 65, 821-838.

Rosenbaum, P. and D. Rubin, 1993, The Central Role of the Propensity Score in Observational Studies for Causal Effects, Biometrika, 70, 41-55.

Rosenzweig, M. and K. Wolpin, 1980, Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment, Econometrica, 48, 227-240.

Rosenzweig, M. and K. Wolpin, 2000, Natural "Natural Experiments" in Economics, Journal of Economic Literature, 38, 827-874.

Rubin, D., 1974, Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, Journal of Educational Psychology, 66, 688-701.

Rubinfeld, D., 1985, Econometrics in the Courtroom, Columbia Law Review, 85, 1048-1119.

Stinchcombe, M. and H. White, 1998, Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative, Econometric Theory, 14, 295-324.

Szymanski, S., 2001, Income Inequality, Competitive Balance and the Attractiveness of Team Sports: Some Evidence and a Natural Experiment from English Soccer, Economic Journal, 111, F69-84.

Taylor, C. and D. Hosken, 2004, The Economic Effects of the Marathon-Ashland Joint Venture: The Importance of Industry Supply Shocks and Vertical Market Structure, FTC Working Paper 270.

Treble, J., 2003, Intertemporal Substitution of Effort: Some Empirical Evidence, Economica, 70, 579-595.

United States General Accounting Office, 2004, Energy Markets: Effects of the Mergers and Market Concentration in the U.S. Petroleum Industry, Report to the Ranking Minority Member, Permanent Subcommittee on Investigations, Committee on Governmental Affairs, U.S. Senate, GAO-04-96.

Vita, M. and S. Sacher, 2001, The Competitive Effects of Not-For-Profit Hospital Mergers: A Case Study, The Journal of Industrial Economics, 49, 63-84.

Wald, A., 1943, Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large, Transactions of the American Mathematical Society, 54, 426-482.

White, H., 1990, Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings, Neural Networks, 3, 535-549

White, H., 1994, Estimation, Inference, and Specification Analysis. New York: Cambridge University Press.

White, H., 2001, Asymptotic Theory for Econometricians. New York: Academic Press.

White, H., 2004, Approximate Nonlinear Forecasting Methods, forthcoming in G. Elliott, A. Timmermann, and C. Granger (eds.) Handbook of Economic Forecasting, Amsterdam: North Holland.

Zveglich, J. and Y. Rodgers, 2003, The Impact of Protective Measures for Female Workers, Journal of Labor Economics, 21, 533-556.