Time-Series Estimation of the Effects of Natural Experiments

Halbert White
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508

June 11, 2005

Abstract This paper builds on the labor econometrics and classical treatment effects literatures to provide a framework supporting causal concepts and methods for estimating effects of natural experiments operating over time in an explicitly dynamic time-series context. We examine conditions for the construction of covariates instrumental in identifying effects of interest that lead to new tests for unconfoundedness, a key condition for the identification of causal effects that we link to the concept of Granger non-causality. Our new tests for unconfoundedness are useful in both cross-section and dynamic time-series settings.

Acknowledgments: The author is grateful for the comments and suggestions of the editor, two anonymous referees, Douglas Bernheim, Karim Chalak, Xiaohong Chen, Clive Granger, Jerry Hausman, David Hendry, Jinyong Han, Kei Hirano, Paul Johnson, Qi Li, Robert Marshall, Jack Porter, Chris Stomberg, Timo Teräsvirta, Aman Ullah, 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. The author thanks Michael J. Bacci and Tolga Cenesizoglu for their invaluable assistance with the preparation of this manuscript

JEL Classification Numbers: B41, C22, C32, C53, L40.

1 Introduction

Two areas of long-standing interest to Clive Granger are time series and causality. Despite the wide application of and considerable insight provided by Clive's seminal investigations into what is now known as "Granger Causality" (Granger, 1969), notions of causality for economic time series are still not settled, in relative contrast to the situation in labor economics, for example. There, concepts and methods from the statistical treatment effects literature have been widely adopted and further developed as reliable means for estimating causal effects of interest in cross-section data.

Recognizing this contrast, Bauwens, Boswijk, and Urbain recently organized an EC^2 conference on "Causality and Exogeneity in Economics," expressly designed to foster interaction between researchers approaching these issues from different perspectives, especially the time-series and cross-section viewpoints. A selection of the papers presented will shortly appear as a special issue of the *Journal of Econometrics* (Bauwens, Boswijk, and Urbain, 2005). The volume represents a significant collection of work by outstanding researchers, but it also provides a clear demonstration of the gap that still exists in attempting to understand causality in the different branches of econometrics.

Recent work of Angrist and Kuersteiner (2004) has begun to bridge this gap by introducing treatment effect-based causality tests suitable for assessing the effectiveness of policy interventions in dynamic time-series contexts. This work draws on methods for the estimation of treatment effects in longitudinal/panel data developed by Robins, Greenland, and Hu (1999). In particular, Angrist and Kuersteiner develop methods for testing for the presence of causal effects associated with recurring interventions, such as the monetary policy interventions studied by Romer and Romer (1989); and they relate their new tests to the classical Granger- and Sims-causality tests (Sims, 1972).

Although Angrist and Kuersteiner provide tests for the presence of causal effects of interventions, they do so indirectly by making use of the "policy propensity score," rather than by directly estimating the effect of the intervention. Our purpose here is to begin to develop methods for directly estimating causal effects of interventions in dynamic time-series contexts, further bridging the gap between the time-series and cross-section perspectives by extending to a dynamic time-series context causal concepts and methods for estimating causal effects developed over the last twenty years or so largely in the labor econometrics and statistical treatment effects literatures. We draw heavily on key contributions in the labor econometrics literature by Heckman, Ichimura, and Todd (1997,1998), Hahn (1998), Hirano and Imbens (2001), and Hirano, Imbens, and Ridder (2003), among others, which extend seminal work in the treatment effects literature by Rubin (1974) and Rosenbaum and Rubin

(1983).

Although the framework developed here is relevant for estimating the effects of recurring interventions, our focus will be on estimating the effects of "natural experiments" operating in time, that is, specific non-recurring interventions that occur over some more or less sustained period of time. Such natural experiments are of particular interest in antitrust analysis, where, for example, interest frequently attaches to the price effects of a particular cartel that may have operated over a period of years or to the price effects of a given merger or acquisition. Other relevant examples of the natural experiments treated here are the implementation of a specific economic or social policy, or the deployment of a new technology. We thus build on work pioneered by Angrist (1998) analyzing the economic effects of natural experiments.

In addition to providing concepts and methods permitting the measurement of the effects of natural experiments operating over time, we make other specific contributions, relevant not only in time-series applications, but also in cross-section settings. Specifically, we examine conditions for the construction of covariates effective in identifying effects of interest that lead to new tests for unconfoundedness, a key condition for the identification of causal effects.

This paper is organized as follows. Section 2 provides an explicit formal framework permitting rigorous analysis of causal effects for a non-dynamic but possibly dependent data generating process. In Section 3 we review extant methods for estimating specific effects of interest, with particular attention to significant developments in the labor econometrics and classical treatment effect literatures. This provides the foundation for the extension to a dynamic time-series setting developed in Section 4. There we discuss notions of effect relevant in dynamic time-series contexts and broadly sketch strategies for estimation of the effects of natural experiments of interest here. Section 5 contains a deeper discussion of issues arising in the selection of covariates, together with a new test for unconfoundedness appropriate either in cross-section or dynamic time-series settings. Concluding remarks appear in Section 6. A mathematical appendix contains proofs of formal results.

2 The Data Generating Process and Effects of Interest

We consider a data generating process with explicit causal structure in order to rigorously define effects of interest and to clearly specify formal conditions permitting identification of these effects.

For concreteness, we operate within the formal causal framework of a settable system, as defined by White (2005b, ch. 3.8). This framework blends concepts explicit and implicit in the classical

Cowles Foundation approach to econometric modeling originated by Tinbergen, Frisch, Koopmans, and Haavelmo, among others (see Goldberger, 1972, 1991) with concepts developed in recent work of Pearl (1988, 1993a, 1993b, 1995, 2000). It is closely related to the intervention directed acyclic graphs (DAGs) of Dawid (2002). The result is an explicitly causal formal framework that corresponds closely to the reduced form equations widely used in econometric practice. Here we briefly summarize the key ideas.

Fundamental to this framework is the notion of settable variables. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\{\mathcal{X}_1, \mathcal{X}_2, \ldots\}$ be a countable collection of mappings $\mathcal{X}_j : \{0,1\} \times \Omega \to \mathbb{R}$ such that

$$\mathcal{X}_j(1, \cdot) = Z_j(\cdot), \qquad j = 1, 2, \ldots$$

for a given collection of random variables $\{Z_j\}$ (whose behavior is governed by P), and for a given measurable mapping c_j

$$\mathcal{X}_j(0, \cdot) = Y_j \equiv c_j(Z_{(j)}), \qquad j = 1, 2, \ldots,$$

where $Z_{(j)} \equiv Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots$. We call the functions c_j *response functions.* We refer to $\mathcal{X}_j(1, \cdot) = Z_j$ as the "setting" of $\mathcal{X}_j$, corresponding to an intervention in Dawid's (2002) intervention DAGs. We refer to $\mathcal{X}_j(0, \cdot) = Y_j$ as the "response" of $\mathcal{X}_j$. We also call $\mathcal{X}_j(0, \cdot) = Y_j$ a "dependent variable." The functions $\mathcal{X}_j$ just defined are *settable variables.*

The pair $\{(\Omega, \mathcal{F}, P), \{Z_j, c_j, \mathcal{X}_j\}\}$ is a *settable system.* We may also simply refer to $\{\mathcal{X}_j\}$ as a settable system.

The key idea is that the response of $\mathcal{X}_j$ is entirely determined by the other variables of the system, set in a manner governed by P. Not all of the other variables may matter in determining the response. If for a given $i \neq j$, $c_j(z_{(j)})$ defines a function constant in z_i, then we say $\mathcal{X}_i$ *does not cause* $\mathcal{X}_j$ with respect to the underlying settable system, and we write $\mathcal{X}_i \not\leadsto \mathcal{X}_j$. Otherwise, we say that $\mathcal{X}_i$ *causes* $\mathcal{X}_j$ and write $\mathcal{X}_i \leadsto \mathcal{X}_j$.

The subject of causality is vast and deep, with a history spanning millennia, so to provide relevant background, it is feasible here only to point to appropriate sources. One reference providing an extensive survey particularly accessible to and relevant for economists is Hoover's (2001) *Causality in Macroeconomics.* Given the immensity of the subject, it would be farcical to view the settable system framework as the definitive way to analyze causality. Rather, we view it as an explicit, parsimonious, and convenient formal framework that matches well with economists' intuitions about causation and that readily lends itself to useful extensions, especially to dynamic contexts.

The response functions c_j provide the basis for a meaningful definition of the concept "effect." For example, when the derivative exists, the *ceteris paribus marginal effect* of $\mathcal{X}_i$ on $\mathcal{X}_j$ can be defined as

$$D_i c_j(z_{(j)}) = \partial c_j(z_{(j)})/\partial z_i.$$

Generally it will not be possible to observe every variable of the settable system. This is accommodated by considering certain average notions of effect, discussed below.

With these concepts, we can specify the data generating process (DGP) as follows:

Assumption A.1 *Let $(\mathcal{Y}, \mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ be a settable system such that:* (i) *$\mathcal{Y}$ is scalar-valued, $\mathcal{D}$ is $\{0,1\}$-valued, $\tilde{\mathcal{Z}}$ is $\mathbb{R}^k$-valued, $k \in \mathbb{N}$, and $\ddot{\mathcal{Z}}$ is $\mathbb{R}^\infty$-valued, $\mathbb{R}^\infty \equiv \mathbb{R} \times \mathbb{R} \times \ldots$*; (ii) *$\mathcal{Y} \not\rightsquigarrow (\mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ and $\mathcal{D} \not\rightsquigarrow (\tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$*; (iii) *$(\mathcal{D}, \mathcal{Z}, \ddot{\mathcal{Z}})$ generates random variables $\{(D_t, \tilde{Z}_t, \ddot{Z}_t)\}$ that determine the responses $\{Y_t\}$ of $\mathcal{Y}$ according to*

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

where c is an unknown measurable scalar-valued function; (iv) *for $i = 0, 1$, define $\mathcal{T}_i \equiv \{t \in \mathbb{N} : D_t = i\}$ and assume for all $t \in \mathcal{T}_i$ that $\tilde{Z}_t$ has joint distribution $\tilde{H}_i$, and the conditional distribution of $\ddot{Z}_t$ given $\tilde{Z}_t = \tilde{z}$ is $\tilde{G}_i(\cdot|\tilde{z})$*; (v) *the realizations of Y_t, D_t, and $\tilde{Z}_t$ are observed, whereas those of $\ddot{Z}_t$ are not observed.*

According to A.1, the observed response Y_t is generated by an observed binary "treatment" D_t and potential causes $\tilde{Z}_t$ and $\ddot{Z}_t$, of which $\tilde{Z}_t$ is observed and $\ddot{Z}_t$ is not. By admitting "potential" causes in $\tilde{Z}_t$, we permit certain elements of $\tilde{Z}_t$ not to matter for the response of $\mathcal{Y}$. As a useful convention, we view all elements of $\ddot{\mathcal{Z}}$ as truly causal for $\mathcal{Y}$, that is, $\ddot{\mathcal{Z}} \rightsquigarrow \mathcal{Y}$. Our statement of A.1 thus leaves implicit other variables of the settable system that are not presently of direct interest, such as unobserved variables not causing $\mathcal{Y}$ or variables caused by $\mathcal{Y}$. Later we explicitly identify certain additional relevant variables of the settable system.

By requiring $\mathcal{Y} \not\rightsquigarrow (\mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, we rule out "endogenous" right hand side variables. The requirement that treatment $\mathcal{D}$ does not cause $\tilde{\mathcal{Z}}$ or $\ddot{\mathcal{Z}}$ plays a key role in properly isolating the effects of interest. This corresponds to avoiding the inclusion of "outcome" variables when modeling causal response, as discussed by Angrist and Krueger (1999, p. 1292). The adverse consequences of not enforcing this requirement are discussed extensively by Rosenbaum (1984).

The settable variables $(\mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ generate $(D_t, \tilde{Z}_t, \ddot{Z}_t)$ either as responses or settings. The imposed causal structure makes it unnecessary to specify explicitly which role (setting or response) these settable variables play, and it may be convenient to view specific variables either as settings or as responses to deeper causal structures, depending on the context. The sequence $\{(D_t, \tilde{Z}_t, \ddot{Z}_t)\}$ is generated by fixing the roles of $(\mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ and from them generating repeated draws $(D_t, \tilde{Z}_t, \ddot{Z}_t)$, $t = 1, 2, \ldots$ whose joint behavior is governed by P. In cross-section settings, P generates independent draws; however, independence is not necessary. Here P can permit arbitrary dependence across observations and in particular temporal dependence.

By specifying that the response is entirely determined by $(D_t, \tilde{Z}_t, \ddot{Z}_t)$ where $\mathcal{D} \rightsquigarrow\!\!| \,(\tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, we ensure that $(\tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ is a set of *sufficient concomitants* in the terminology of Dawid (2000).

The response function c embodies the effects of interest, specifically the effect of the natural experiment or treatment, D_t. Assumption A.1 imposes minimal structure on the response. In particular, it does not impose linearity or separability of the response with respect to observed and unobserved variables, as is often done in economics (e.g., see Heckman, Ichimura, and Todd (1998), for discussion).

We refer to $\mathcal{T}_0$ and $\mathcal{T}_1$ as "regimes" 0 and 1, respectively. In the cases of interest here, $\mathcal{T}_0$ and $\mathcal{T}_1$ may be sequential. For example, if the natural experiment is the formation of a cartel, then $\mathcal{T}_0$ may represent the "pre-cartel" regime and $\mathcal{T}_1$ the cartel regime. On the other hand, one may have only cartel and "post-cartel" data, in which case $\mathcal{T}_1$ precedes $\mathcal{T}_0$. The regimes need not be sequential, however. One may have data both pre- and post-cartel, in which case $\mathcal{T}_0$ represents the "ex-cartel" regime that both precedes and follows the cartel regime.

The fact that $\mathcal{T}_0$ and $\mathcal{T}_1$ may themselves contain sequences of observations in a time-series context (where the sequence of observations matters) is an important aspect of the present setup that distinguishes it from the cross-section setting in which labor economics has developed its methods for estimating causal effects. In particular, this has important implications for the timing of the measurement of the right-hand side variables. In cross-sections, it is usually assumed that covariates are measured prior to 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 ...

In the present setting, however, observations on potential causes are almost always made after the onset of the treatment: for example, we observe cost and demand shifters that matter for price determination after a merger or formation of a cartel. The requirement that $\mathcal{D} \not\Rightarrow (\tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ thus has particular importance here: for example, in antitrust analysis we must exclude from consideration any potential causes that would be affected by the merger or cartel.

A particularly notable consequence of this requirement is that Assumption A.1 rules out explicit inclusion of lagged dependent variables among the elements of $\tilde{Z}_t$ and $\ddot{Z}_t$. In a cartel or merger application, this essentially requires imposing the assumption that any relevant market dynamics work themselves out within the observation period, so that the observations comprise a sequence of equilibria; but this sidesteps important cases of interest involving dynamics. In Section 4 we extend the present framework to handle dynamics. For now, however, we work with a more restrictive framework, adequate for the extant treatment effects and labor econometrics literatures, in order to readily grasp the roles played by its various components. This will also provide the needed foundation for our later dynamic framework.

Assumption A.1 explicitly permits the distributions of $\tilde{Z}_t$ and of $\ddot{Z}_t$ given $\tilde{Z}_t$ to differ between regimes. If economists could conduct randomized experiments, this possibility would be unnecessary, as randomization acts to help ensure $\tilde{H}_0 = \tilde{H}_1$ and $\tilde{G}_0 = \tilde{G}_1$. In the present context, we may well expect to have these distributions differ. For example, distributions of cost and demand shifters may well differ between cartel and ex-cartel regimes for reasons causally unrelated to the cartel. Methods for analyzing non-randomized experiments pioneered by Rubin (1974) and Rosenbaum and Rubin (1983) and in labor economics (e.g. Heckman, Ichimura, and Todd, 1997, 1998; Hahn, 1998; Hirano and Imbens, 2001; Hirano, Imbens, and Ridder, 2003) permit one to avoid confounding the effect of interest with shifts in the distribution of other causes.

Assumption A.1 permits us to formally define the effects of interest here. The effect of the natural experiment given $(\tilde{z}, \ddot{z})$ is

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

where we write $c_i(\tilde{z}, \ddot{z}) \equiv c(i, \tilde{z}, \ddot{z})$ for $i = 0, 1$. This is a standard definition of a treatment effect (see, e.g. Heckman, Ichimura, and Todd, 1998). This effect is unobservable, however, because $c_1(\tilde{z}, \ddot{z})$ and $c_0(\tilde{z}, \ddot{z})$ are not simultaneously observable: we can usually observe only one of these quantities.

Nevertheless, averaging leads to effects that can be estimated, under suitable assumptions. The effect that we focus on in this section is the *average effect of treatment on the treated*:

$$\begin{aligned}\Delta_1^* &\equiv \int \Delta(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z} \,|\, \tilde{z})\, d\tilde{H}_1(\tilde{z}) \\ &= E(Y_t^1|D_t = 1) - E(Y_t^0|D_t = 1),\end{aligned}$$

where $Y_t^1 \equiv c_1(\tilde{Z}_t, \ddot{Z}_t)$, $Y_t^0 \equiv c_0(\tilde{Z}_t, \ddot{Z}_t)$ (cf. Rubin, 1974).

In the cartel or merger examples, Δ_1^* represents the average effect of the cartel or merger, averaging over underlying market conditions in that regime. For example, this is the difference between the average price in the cartel regime,

$$\mu_1 \equiv E(Y_t^1|D_t = 1) = \int c_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, d\tilde{H}_1(\tilde{z})$$

and the average "but-for" price, the counter-factual price that would have prevailed under the same market conditions, but in the absence of the cartel,

$$\mu_{01} \equiv E(Y_t^0|D_t = 1) = \int c_0(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z} \,|\, \tilde{z})\, d\tilde{H}_1(\tilde{z}).$$

3 Estimating Effects of Interest

A simple approach to attempting to estimate the effects of a natural experiment is to apply ordinary least squares (OLS) to the "dummy variable" model

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

where $Z_t \equiv (1, \tilde{Z}_t)'$. The OLS estimator is

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

where D is the $T \times 1$ vector with elements D_t, Z is the $T \times k$ matrix with rows Z_t' and Y is the $T \times 1$ vector with elements Y_t. The traditional textbook interpretation of OLS (e.g. Greene, 1993, pp. 231-232) holds that $\hat{\alpha}$ estimates the ceteris paribus effect of the natural experiment, and $\hat{\beta}$ estimates the ceteris paribus effects of all the other included regressors.

One area where the dummy variable method is widely used is in antitrust, where, in allegations of price fixing, interest attaches to the price impact of a cartel or where, in the regulation of mergers and acquisitions, interest attaches to the price impact of a given merger or acquisition. The articles of Fisher (1980) and Rubinfeld (1985) are classics in this area that have had a lasting impact in sustaining the use of the dummy variable method. The importance and common use of this method in antitrust have been noted as recently as 2003 by Higgins and Johnson (2003), and its use continues. Recent examples in the analysis of mergers are Vita and Sacher (2001), United States General Accounting Office (2004), and Taylor and Hosken (2004).

In contrast, modern labor economics has largely abandoned the use of dummy variable models in favor of more flexible methods. In their important survey, Angrist and Krueger (1999) provide an extensive analysis of a wide variety of approaches to estimating effects of interest in labor economics, detailing the advantages and pitfalls of the various methods. In particular, Angrist and Krueger discuss pitfalls associated with use of dummy variable methods.

As a complement to Angrist and Krueger's discussion and to explicitly expose the various sources of the pitfalls of the dummy variable approach, in Proposition 3.1 we offer a decomposition of the apparent effect estimated by $\hat{\alpha}$.

For this, we make use of the following notation and assumption. Let $\hat{p}_1 = T_1/T$, where T_1 is the number of regime one observations, and write $T_0 \equiv T - T_1$,

$$\hat{M}_0 \equiv T_0^{-1} \sum_{t=1}^{T} (1 - D_t) Z_t Z_t' \qquad \hat{M}_1 \equiv T_1^{-1} \sum_{t=1}^{T} D_t Z_t Z_t'$$
$$\hat{L}_0 \equiv T_0^{-1} \sum_{t=1}^{T} (1 - D_t) Z_t Y_t \qquad \hat{L}_1 \equiv T_1^{-1} \sum_{t=1}^{T} D_t Z_t Y_t.$$

Assumption A.2

(a) $\hat{p}_1 \xrightarrow{p} p_1, \quad 0 < p_1 < 1.$

(b) (i) $\hat{M}_0 \xrightarrow{p} M_0 \equiv \int zz' d\tilde{H}_0(\tilde{z}) < \infty, \quad \det M_0 > 0$

(ii) $\hat{M}_1 \xrightarrow{p} M_1 \equiv \int zz' d\tilde{H}_1(\tilde{z}) < \infty, \quad \det M_1 > 0.$

(c) (i) $\hat{L}_0 \xrightarrow{p} L_0 \equiv \int zc_0(\tilde{z}, \ddot{z})\, d\tilde{G}_0(\ddot{z}|\tilde{z})\, d\tilde{H}_0(\tilde{z}) < \infty$

(ii) $\hat{L}_1 \xrightarrow{p} L_1 \equiv \int zc_1(\tilde{z}, \ddot{z})\, d\tilde{G}_1(\ddot{z}|\tilde{z})\, d\tilde{H}_1(\tilde{z}) < \infty$

Proposition 3.1 *Suppose Assumptions A.1 and A.2 hold and that in addition* $L_{01} \equiv \int z c_0(\tilde{z}, \ddot{z}) \, d\tilde{G}_1(\ddot{z}|\tilde{z}) \; d\tilde{H}_1(\tilde{z}) < \infty$. *Then* $\hat{\alpha} \xrightarrow{p} \alpha^*$, *where*

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

where $m_0 \equiv \int z \; d\tilde{H}_0(\tilde{z})$, $m_1 \equiv \int z \, d\tilde{H}_1(\tilde{z})$, $\beta_0^* \equiv M_0^{-1} L_0$, $\beta_1^* \equiv M_1^{-1} L_1$, $\beta_{01}^* \equiv M_1^{-1} L_{01}$. S is *the selection matrix that selects the non-constant elements of* z $(\tilde{z}' = Sz)$, *and* $\tilde{M} \equiv (1 - p_1)\tilde{M}_0 + p_1 \tilde{M}_1$, $\tilde{M}_0 \equiv S(M_0 - m_0 m_0')S'$, $\tilde{M}_1 \equiv S(M_1 - m_1 m_1')S'$.

This result is useful for two reasons. First, it reveals conditions under which $\alpha^* = \Delta_1^*$, so that OLS indeed estimates the effect of interest. From the second expression it follows immediately that $\alpha^* = \Delta_1^*$ provided that $\tilde{H}_0 = \tilde{H}_1$ and $\tilde{G}_0 = \tilde{G}_1$, regardless of whether the dummy variable model is correctly specified. From the first expression, a little analysis shows that if the dummy variable model is correctly specified (i.e. $c_0(\tilde{z}, \ddot{z}) = z'b^* + u(\ddot{z})$, $c_1(\tilde{z}, \ddot{z}) = \delta^* + c_0(\tilde{z}, \ddot{z})$ with $\int u(\ddot{z})\, d\tilde{G}_0(\ddot{z} \,|\, \tilde{z}) = 0$), and if also $\tilde{G}_0 = \tilde{G}_1$, then $\alpha^* = \Delta_1^*$.

Second, this result shows explicitly what happens when both of these two sufficient conditions for $\alpha^* = \Delta_1^*$ fail. This is important because neither sufficient condition is particularly plausible in applications. Proposition 3.1 offers insight into how the apparent effect α^* can be manipulated by selection of regressors $\tilde{Z}_t$ to achieve whatever value one might like. This decomposition can be particularly useful when one is attempting to understand which aspects of the data generate a given apparent effect.

Thus, not only does the dummy variable method have the drawback that it is consistent for the effect of interest only under quite stringent conditions, but it has the further drawback that it is subject to manipulation in their absence. Analysis of effects of interest in antitrust or other areas of public policy is therefore not well served by use of the dummy variable method.

Modern labor econometrics has drawn on and considerably extended methods from the treatment effects literature (e.g. Rubin, 1974; Rosenbaum and Rubin, 1983) to obtain less problematic estimators for causal effects of interest. Key contributions are those of Heckman, Ichimura, and Todd (1997, 1998), Hahn (1998), Hirano and Imbens (2001), and Hirano, Imbens, and Ridder (2003), among others. These methods have been developed explicitly for cross-section applications, but, as we now discuss, with suitable adaptation and extension, they can have general utility, especially in areas involving time-series or panel data, such as antitrust, finance, and macroeconomics.

Space is not available for a survey of the treatment effects and labor econometrics literatures. Instead we illustrate main ideas with a few key papers.

The simple idea underlying the various estimators of $\Delta_1^* = \mu_1 - \mu_{01}$ is to estimate μ_1 and μ_{01} directly and as (jointly) efficiently as possible. It is easy to estimate μ_1; an obvious consistent estimator is the regime 1 average,

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

Given $\hat{\mu}_{01}$ consistent for $\mu_{01} \equiv E(Y_t^0 | D_t = 1)$, a consistent estimator for Δ_1^* is

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

Finding a consistent estimator $\hat{\mu}_{01}$ poses a challenge, however, as Y_t^0 cannot be observed when $D_t = 1$.

In a classic of the treatment effects literature, Rubin (1974) resolved this challenge by introducing the *unconfoundedness* condition: (Y_t^0, Y_t^1) is independent of D_t given X_t, a vector of observable "covariates" not causally impacted by treatment. (Note that X_t is conceptually distinct from $\tilde{Z}_t$. We discuss the relation between X_t and $\tilde{Z}_t$ at length below.) Expressed in the notation of conditional independence (see Dawid, 1979), the unconfoundedness condition is

$$(Y_t^0, Y_t^1) \perp D_t \mid X_t.$$

This condition is also known as "selection on observables" (see Barnow, Cain, and Goldberger, 1981). Below, we relate this condition to the concept of Granger (1969) non-causality.

With unconfoundedness, we have

$$\begin{aligned}\tilde{\mu}_{01}(X_t) &\equiv E(Y_t^0 \mid D_t = 1, X_t) \\ &= E(Y_t^0 \mid D_t = 0, X_t) \\ &\equiv \tilde{\mu}_0(X_t).\end{aligned}$$

Given $\tilde{\mu}_0$, a consistent estimate of μ_{01} can be obtained as

$$\hat{\mu}_{01} = T_1^{-1} \sum_{t \in \mathcal{T}_1} \tilde{\mu}_0(X_t).$$

Put succinctly, many of the advances in labor econometrics exploit unconfoundedness to obtain estimates of $\tilde{\mu}_0$ (hence $\tilde{\mu}_{01}$) and thence μ_{01}.

Rosenbaum and Rubin (1983) showed that unconfoundedness implies

$$(Y_t^0, Y_t^1) \perp D_t \mid p(X_t),$$

where $p(X_t) \equiv P[D_t = 1 \mid X_t]$ is the "propensity score": the probability of treatment given the covariates. Advances in labor econometrics have also exploited properties of the propensity score.

A useful sampling of these advances is provided by the papers of Hahn (1998), Hirano, Imbens, and Ridder (2003) (HIR), and Hirano and Imbens (2001) (HI). We now briefly sketch their features salient here.

Hahn (1998) exploited unconfoundedness to estimate μ_{01} as

$$\hat{\mu}_{01}^H = T_1^{-1} \sum_{t=1}^{T} D_t \hat{\beta}_0(X_t),$$

where $\hat{\beta}_0(X_t) \equiv \hat{E}[(1 - D_t)Y_t | X_t]/(1 - \hat{E}[D_t | X_t])$, 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 $\hat{\mu}_1^H = T_1^{-1} \sum_{t=1}^{T} D_t \hat{\beta}_1(X_t)$, where $\hat{\beta}_1(X_t) \equiv \hat{E}[D_t Y_t \mid X_t]/\hat{E}(\hat{D}_t \mid X_t)$. Thus Hahn's estimator for the effect of interest is

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

This estimator requires nonparametric estimation of $E[(1-D_t)Y_t|X_t]$, $E[D_tY_t|X_t]$, and $E(D_t|X_t)$. Hahn proposes nonparametric series estimators (polynomials in X_t) and, in his Theorem 6, provides conditions ensuring $\hat{\Delta}_1^H$ is consistent for Δ_1^*, asymptotically normal, and attains the semi-parametric efficiency bound.

Hirano, Imbens, and Ridder (2003) (HIR) note that the burden of nonparametric estimation posed by $\hat{\Delta}_1^H$ can be significantly reduced by use of the propensity score. They propose an estimator of Δ_1^* that can be written

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

where $\hat{\mu}_1$ is the regime 1 sample mean as above, and

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

with $\hat{p}$ a nonparametric estimator of the propensity score. The HIR estimator eliminates the need to estimate $E[(1-D_t)Y_t \mid X_t]$ and $E[D_tY_t \mid X_t]$. HIR propose a logistic series estimator for constructing $\hat{p}$, and in their Theorem 5 provide conditions under which $\hat{\Delta}_1^{HIR}$ is consistent for Δ_1^*, asymptotically normal, and also attains the semiparametric efficiency bound.

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

$$\mu_{00}(h) \equiv \int \tilde{\mu}_0(x) h(x) \, dH_0(x),$$

with h chosen as $dH_1/\,dH_0$, as then $\mu_{00}(dH_1/\,dH_0) = \mu_{01}$. HIR's estimator approximates $dH_1/\,dH_0$ 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 in regime 0. (Note that care must be taken to ensure dH_1/dH_0 is well defined; a sufficient condition is that H_1 and H_0 have common support.)

Hirano and Imbens (2001) (HI) draw on the work of HIR and propose an estimator that can be written

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

with

$$\hat{\mu}_{01}^{HI} \equiv \hat{m}'_{\psi,1}\hat{\beta}_{\psi,0}^{HI} \equiv T_1^{-1}\sum_{t=1}^{T} D_t X'_{\psi t}\hat{\beta}_{\psi,0}^{HI},$$

where $\hat{\beta}_{\psi,0}^{HI}$ is estimated from a regime 0 regression and $X_{\psi t} \equiv (1, \psi_1(X_t), \ldots, \psi_{q-1}(X_t))$ is a vector of transformations of the covariates. The ψ_j's are known functions; for example, they may implement the polynomials of Hahn (1998) and HIR.

The regime 0 estimator $\hat{\beta}_{\psi,0}^{HI}$ is the weighted least squares (WLS) estimator

$$\hat{\beta}_{\psi,0}^{HI} \equiv (\sum_{t=1}^{T} \hat{\omega}_t(1-D_t)X_{\psi t}X'_{\psi t})^{-1}\sum_{t=1}^{T} \hat{\omega}_t(1-D_t)X_{\psi t}Y_t,$$

where $\hat{\omega}_t$ is computed from a flexible parametric estimator $\hat{p}$ of the propensity score as $\hat{\omega}_t \equiv \hat{p}(X_t)/(1-\hat{p}(X_t))$. The role of the weights is not to achieve efficiency á la GLS, but rather to reweight the regime 0 covariate distribution to that of regime 1, similar to HIR (cf. Horvitz and Thompson, 1952). The HIR estimator corresponds to the special case in which $X_{\psi t} \equiv 1$.

As HI note, $\hat{\Delta}_1^{HI}$ can be conveniently obtained as the WLS estimator $\hat{\alpha}$ of α in the regression

$$Y_t = D_t\alpha + X'_{\psi t}\beta + D_t(\tilde{X}_{\psi t} - \tilde{m}_{\psi,1})'\gamma + \varepsilon_t,$$

where $\hat{\omega}_t = \hat{p}(X_t)/(1-\hat{p}(X_t))$ for regime 0 and $\hat{\omega}_t = 1$ for regime 1, and $\tilde{m}_{\psi,1}$ is the regime 1 average of $\tilde{X}_{\psi,t} \equiv (\psi_1(X_t), \ldots, \psi_{q-1}(X_t))$.

An important aspect of regressions of this sort is that although the estimator $\hat{\alpha}$ is consistent for the effect of interest Δ_1^*, the other estimated regression coefficients generally do not have the textbook ceteris paribus causal interpretation. (See Barnow, Cain, and Goldberger (1981) and Angrist and Krueger, (1999, pp. 1290).) Instead, the slope coefficient estimator (here $\hat{\beta}_{\psi,0}^{HI}$) has a *predictive* interpretation, in that $X'_{\psi t}\hat{\beta}_{\psi,0}^{HI}$ provides a mean-squared-error-optimal approximation to the predictor $\tilde{\mu}_0(X_t)$. The lack of a ceteris paribus interpretation arises because there is no requirement or guarantee that the covariates X_t are accurate measurements of observed true causes ($\tilde{Z}_t$) independent of unobserved causes ($\ddot{Z}_t$).

A significant consequence of this fact is that the signs and magnitudes of the estimated regression slope coefficients need not and generally will not conform to signs or magnitudes that economic theory might predict for a truly causal relationship. Consequently, the "wrong" signs or magnitudes for the slope coefficient estimators cannot be used as legitimate diagnostics for the validity of the estimated effect of interest.

The optimal prediction property of $X'_{\psi t}\hat{\beta}^{HI}_{\psi,0}$ also provides useful insight into the interpretation of $\hat{\Delta}_1^{HI}$ and similar estimators: the estimated effect of treatment on the treated is the average of the difference between actual outcomes for regime 1 and predicted values based on the regime 0 predictor $\tilde{\mu}_0$ applied to the economic conditions of regime 1. This interpretation is especially powerful, as it extends directly to the dynamic case, as we see shortly.

Although HI do not formally treat their estimator as nonparametric, they note the possibility of allowing their models for the response and the propensity score to become more flexible ($q \to \infty$) as $n \to \infty$. By applying results on sieve estimation (see Chen (2005) for a comprehensive survey), fully nonparametric results for the HI estimator can be obtained under conditions admitting time-series without dynamics. The results of Chen (2005) also apply to establish the asymptotic properties of $\hat{\Delta}_1^{H}$ and $\hat{\Delta}_1^{HIR}$ in such time-series contexts.

The work of Hahn (1998), HIR, HI, and other advances in the labor econometrics literature thus provides methods for estimating treatment effects that can be carried into a simple time-series setting just by weakening the i.i.d assumption appropriate to cross-section data to mild dependence conditions that can nevertheless deliver suitable laws of large numbers and central limit results.

We note in passing that certain aspects of these estimators are not necessarily well-advised in applications. Polynomials do not accommodate categorical variables and generally do not yield well-behaved sieve estimators. Further, the model selection methods proposed by HI may suffer from pre-testing problems. Li, Racine, and Wooldridge (2004) (LRW) provide some effective alternative methods. In applications to natural experiments, certain properties of the propensity score require the exercise of particular care, as we discuss further in Section 4. Regardless of any limitations, however, these articles provide a valuable body of concepts and approaches for exploiting unconfoundedness and the propensity score to estimate effects of interest.

In the labor econometrics literature, unconfoundedness is typically simply assumed. There is nothing in the framework supplied by Assumption A.1, however, that can plausibly justify unconfoundedness. It is possible to verify unconfoundedness with $X_t = \tilde{Z}_t$ if $\tilde{G}_0 = \tilde{G}_1$, but there is no reason *a priori* to believe that this condition holds in the absence of randomization.

A resolution to this challenge is to find observable variables, say W_t, that can serve as proxies for the unobserved $\ddot{Z}_t'$s, as we discuss in detail in Section 5; such variables are often available as error-laden measurements of $\ddot{Z}_t$. Formally, we seek observable settable variables $\mathcal{W}$ statistically related to

$\ddot{\mathcal{Z}}$. The elements of $\mathcal{W}$ are not causes of $\mathcal{Y}$; if they were, they would already be included in $\tilde{\mathcal{Z}}$. In this sense, they are "irrelevant regressors" in the usual econometric jargon. Further, because W_t must be statistically related to the omitted $\ddot{Z}_t$, W_t is "correlated with the regression error" in the usual jargon. The proxies W_t cannot therefore be instrumental variables in the classical sense, even though it turns out they are indeed instrumental in estimating the effects of interest. Other properties required of $\mathcal{W}$ are that neither $\mathcal{Y}$ nor $\mathcal{D}$ causes $\mathcal{W}$.

We call auxiliary variables $\mathcal{W}$ that are statistically related to $\ddot{\mathcal{Z}}$ and that do not cause $\mathcal{Y}$ "predictive proxies for unobservable causes," or simply "predictive proxies." The word "predictive" is intended to emphasize the fact that the elements of $\mathcal{W}$ are not causally related to $\mathcal{Y}$. If, in addition, neither $\mathcal{Y}$ nor $\mathcal{D}$ causes $\mathcal{W}$, we call $\mathcal{W}$ *valid* predictive proxies.

The instrumental role played by $\mathcal{W}$ is that, as further discussed in Section 5, it can be plausible that with $X_t = (\tilde{Z}_t, W_t)$, we have

$$\ddot{Z}_t \perp D_t \mid X_t.$$

When this condition holds, we call X_t a vector of "proper" covariates, however X_t may have been constructed. When this condition holds with X_t constructed using valid predictive proxies, we say that "conditional independence given predictive proxies" or CIPP holds.

The CIPP condition can be interpreted as the statement that D_t does not Granger-cause $\ddot{Z}_t$ given X_t. (See Florens and Mouchart (1982) and Florens and Fougere (1995)). The spirit of Granger-causality is preserved in that the cause D_t is logically (and temporally, even if not explicit in the time index) prior to the response (if any) represented by $\ddot{Z}_t$. An interesting nuance of CIPP, developed in detail in Section 5, is that X_t need not be logically or temporally prior to $\ddot{Z}_t$, but may instead contain elements that are responses to $\ddot{Z}_t$.

Formally, we have that CIPP implies unconfoundedness:

Proposition 3.2 *If* $\ddot{Z}_t \perp D_t \mid (\tilde{Z}_t, W_t)$, *then* $(c_0(\tilde{Z}_t, \ddot{Z}_t), c_1(\tilde{Z}_t, \ddot{Z}_t)) \perp D_t \quad \mid (\tilde{Z}_t, W_t)$.

Bearing in mind the link between CIPP and Granger non-causality just described and the fact that CIPP implies unconfoundedness, we can now appreciate the link between Granger non-causality and unconfoundedness mentioned earlier.

When CIPP holds, the methods of Hahn, HIR, HI, LRW, or any methods based on unconfoundedness thus deliver informative estimates of effects of interest in time-series settings where the following extension of A.1 holds, together with other suitable regularity conditions.

Assumption B.1 (Data Generating Process) *Let* $(\mathcal{Y}, \mathcal{D}, \mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ *be a settable system such that:* (i) $\mathcal{Y}$ *is scalar-valued,* $\mathcal{D}$ *is* $\{0,1\}$*-valued,* $\mathcal{W}$ *is* $\mathbb{R}^{\ell}$*-valued,* $\ell \in \mathbb{N}$, $\tilde{\mathcal{Z}}$ *is* $\mathbb{R}^{k}$*-valued,* $k \in \mathbb{N}$, *and* $\ddot{\mathcal{Z}}$ *is* $\mathbb{R}^{\infty}$*-valued;* (ii) $Y \not\Leftarrow (\mathcal{D}, \mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, $\mathcal{D} \not\Leftarrow (\mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, *and* $\mathcal{W} \not\Leftarrow \mathcal{Y}$; (iii) $(\mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ *generates random variables* $\{(D_t, \tilde{Z}_t, \ddot{Z}_t)\}$ *that determine the responses* $\{Y_t\}$ *of* $\mathcal{Y}$ *according to*

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

where c is an unknown measurable scalar-valued function; (iv) *for* $i = 0, 1$, *define* $\mathcal{T}_i = \{t \in \mathbb{N} : D_t = i\}$ *and assume for all* $t \in \mathcal{T}_i$ *that* $X_t = (\tilde{Z}_t, W_t)$ *has joint distribution* H_i; *that for all t the conditional distribution of* $\ddot{Z}_t$ *given* $(D_t, X_t) = (d, x)$ *is* $G(\cdot \mid d, x)$; *and that* $\ddot{Z}_t \perp D_t | X_t$; (v) *The realizations of* Y_t, D_t, *and* X_t *are observed, whereas those of* $\ddot{Z}_t$ *are not.*

Assumption B.1 extends A.1 by introducing the observable auxiliary variables $\mathcal{W}$ generating the W_t's that deliver the conditional independence explicitly imposed in (iv).

This framework allows simple time-series dependence, but still does not admit dynamics. Nevertheless, we now have a proper foundation on which to build a framework for estimating the effects of natural experiments in a truly dynamic setting, and we now turn our attention to this task.

4 Introducing Dynamics

The causal structures considered so far permit time-series dependence, but not dynamics generated by the presence of lagged dependent variables in the response function. Such dynamics may play a central role in the behavior of economic phenomena subject to the natural experiments of interest here. The issues involved in treating the dynamic case are extensive, so in the space available here we can only introduce some basic notions and sketch in broad outline methods for estimating effects of interest. We do this by extending the concepts and methods of Sections 2 and 3 in a natural way.

We now work with a *dynamic* settable system. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and now let $\{\mathcal{X}_1, \mathcal{X}_2, \ldots\}$ be a countable collection of mappings $\mathcal{X}_j : \{0,1\} \times \Omega \to \mathbb{R}^{\infty}$. A typical mapping $\mathcal{X}_j$ is now a countable *sequence* of random scalars such that the settings of $\mathcal{X}_j$ are

$$\mathcal{X}_j(1, \cdot) = \{Z_{j1}, Z_{j2}, \ldots\},$$

where $Z_j \equiv \{Z_{jt}\}$ is a given time series of random variables governed by P. To illustrate the way in which we accommodate dynamics, for a given sequence of measurable mappings $c_j = \{c_{jt}\}$ (response functions) and random variables Y_{j0} ("initial values"), let the responses $\mathcal{X}_j(0, \cdot)$ be the sequence $\{Y_{j1}, Y_{j2}, \ldots\}$ defined recursively as

$$Y_{jt} = c_{jt}(Y_j^{t-1}, Z_{(j)}^t), \qquad t = 1, 2, \ldots,$$

where $Y_j^{t-1} \equiv (Y_{j0}, Y_{j1}, \ldots, Y_{jt-1})$ and $Z_{(j)}^t \equiv (Z_{(j)1}, \ldots, Z_{(j)t})$, where $Z_{(j)\tau}$ is the countable vector containing $Z_{i\tau}$, $i \neq j$, $i = 1, 2, \ldots$. The responses or dependent variables $\mathcal{X}_{jt}(0, \cdot)$ thus depend on the history of previous responses and on the history of the previous and current settings. The mappings $\mathcal{X}_j$ just defined are settable *sequences*. The pair $\{(\Omega, \mathcal{F}, P), \{Z_j, c_j, Y_{j0}, \mathcal{X}_j\}\}$ is a *dynamic settable system.*

By recursive substitution, define the *cumulative* response functions $\{c_j^t\}$ as

$$\begin{aligned} Y_{jt} &= c_{jt}(c_{jt-1}(\ldots, c_{j1}(Y_{j0}, Z_{(j)}^1), \ldots, Z_{(j)}^{t-1}), Z_{(j)}^t) \\ &\equiv c_j^t(Y_{j0}, Z_{(j)}^t), \qquad t = 1, 2, \ldots. \end{aligned}$$

We define notions of cause and effect using the c_j^t's. Specifically, for $i \neq j$, if $c_j^t(y_0, z_{(j)}^t)$ defines a function constant in $z_{i\tau_1}, \ldots, z_{i\tau_2}$, $\tau_1 \leq \tau_2$, then we say $\mathcal{X}_{i,\tau_1}^{\tau_2}$ does not cause $\mathcal{X}_{j,t}$ ($\mathcal{X}_{i,\tau_1}^{\tau_2} \not\rightsquigarrow \mathcal{X}_{jt}$). If for all t, $\mathcal{X}_{i,1}^t$ does not cause $\mathcal{X}_{jt}$, then we say $\mathcal{X}_i$ does not cause $\mathcal{X}_j$ ($\mathcal{X}_i \not\rightsquigarrow \mathcal{X}_j$). Otherwise we say $\mathcal{X}_i$ causes $\mathcal{X}_j$ and write $\mathcal{X}_i \rightsquigarrow \mathcal{X}_j$.

Note that future variables cannot cause past variables. Contemporaneous causation is permitted but can be excluded by redefinition of indexes.

Effects are defined similarly to the definitions in Section 2. For example, when the derivative exists, the ceteris paribus marginal effect of $\mathcal{X}_{i\tau}$ on $\mathcal{X}_{jt}$ is

$$D_{i\tau} c_j^t(y_0, z_{(j)}^t) \equiv \partial c_j^t(y_0, z_{(j)}^t) / \partial z_{i\tau}.$$

As before, the fact that not all of the causes can be observed will necessitate consideration of certain expected effects.

At this point we could give a formal statement of the DGP for the dynamic case, analogous to A.1. Nevertheless, for succinctness, we defer a formal statement until we can provide an analog of B.1.

To study natural experiments in a context analogous to that of Assumptions A.1 or B.1, we consider responses $\{Y_t\}$ whose realizations are given by

$$y_t = c^t(y_0, d^t, \tilde{z}^t, \ddot{z}^t), \qquad t = 1, 2, \ldots,$$

where $\tilde{z}^t$ and $\ddot{z}^t$ are realizations of partial histories of observed and unobserved causes, respectively, and d^t is a realization of a partial history of the *dynamic treatment* $d = (d_1, d_2, \ldots)$, a sequence of $\{0,1\}$-valued scalars. Thus, d is the dynamic analog of the treatment of earlier sections. When the dynamic context is understood, we may revert to simply calling d a "treatment."

To study the effects of the natural experiments of interest here, let δ_0 denote a "reference" treatment and let δ_1 denote a "comparison" treatment, $\delta_1 \neq \delta_0$. For concreteness, we take $\delta_0 = (0, 0, \ldots)$ and consider the comparison treatment $\delta_1 = (1, \ldots, 1, 0, 0, \ldots)$. When $\delta_{1t} = 1$ we say that the system is *subject to the natural experiment* or that the "natural experiment operates." As specified, the natural experiment operates continuously for the first T_1 (say) periods and then ceases. We thus call observations 1 through T_1 the *natural experiment.* By convention, the initial value y_0 is not subject to the natural experiment.

Our consideration of reference and comparison treatments is similar to the approach of Robins, Greenland, and Hu (1999) (RGH), who consider the relative outcomes in a panel setting of a reference treatment in which $\delta_{i0} = (0, \ldots, 0)$ and a comparison treatment in which $\delta_{i1} = (1, \ldots, 1)$ for individual i. An important difference between the present time-series context and the panel context is that there is only one available history of observed treatment in the time series, whereas there are multiple histories of observed treatments in panel data. Further, whereas in RGH's panel context it is possible that for no individual i is it true that $D_i = \delta_{i0}$ or $D_i = \delta_{i1}$ (that is, no individual necessarily experiences the reference or comparison treatment), in our time-series context, the observed sample value $D^T(\omega)$ is the realization δ_1^T. That is, the actual outcome $D^T(\omega)$ represents the comparison treatment. Nevertheless, there is a subset of the observed time series representing the reference treatment, as $D_{T_1+1}^T(\omega) = \delta_0^{T-T_1} = (0, \ldots, 0)$. This makes it possible to estimate the behavior of the system outside the operation of the natural experiment.

Although for simplicity and concreteness we explicitly consider the comparison treatment $\delta_1 = (1, \ldots, 1, 0, 0 \ldots)$, the discussion that follows also applies directly to any comparison treatment in which the natural experiment operates over one or more contiguous periods and either does not recur or recurs a few times at most. Examples are comparison treatments of the form $(0, \ldots, 0, 1, \ldots 1, 0, 0, \ldots)$

or $(0, \ldots, 0, 1, 1, \ldots)$. The former comparison treatment, in which the natural experiment occurs between regimes not subject to the natural experiment, is well-suited to the study of the effect of a cartel, whereas the latter comparison treatment, in which the natural experiment operates indefinitely following a regime not subject to its operation, is well suited to the study of the effects of a merger or acquisition. Taking $\delta_1 = (1, \ldots, 1, 0, \ldots)$ permits us to interpret the observation index t conveniently as the number of periods since the start of the natural experiment.

A key distinguishing feature of our current focus is the non-recurrence or minimal recurrence of the natural experiment. This is natural for cartel or merger applications, as well as for the study of the effects of the dissemination of new technologies or the implementation of new laws or social policies, but it is less appropriate for the study of the effects of recurring policy interventions, such as those considered by Romer and Romer (1989) or Angrist and Kuersteiner (2004). Certain of the concepts formalized here are indeed useful in analyzing the effects of these interventions, but the appropriate statistical methods differ from those discussed below. A main reason for this, discussed in more detail shortly, is the nature of the dynamic analog of the propensity score for the natural experiments of interest here.

To examine the effect of the natural experiment, put

$$\begin{aligned} c_0^t(y_0, \tilde{z}^t, \ddot{z}^t) &\equiv c^t(y_0, \delta_0^t, \tilde{z}^t, \ddot{z}^t) \\ c_1^t(y_0, \tilde{z}^t, \ddot{z}^t) &\equiv c^t(y_0, \delta_1^t, \tilde{z}^t, \ddot{z}^t). \end{aligned}$$

For given $y_0, \tilde{z}^t, \ddot{z}^t$, the effect of the natural experiment at time t is

$$\Delta_t(y_0, \tilde{z}^t, \ddot{z}^t) \equiv c_1^t(y_0, \tilde{z}^t, \ddot{z}^t) - c_0^t(y_0, \tilde{z}^t, \ddot{z}^t).$$

The fact that $\ddot{z}^t$ is unobserved necessitates considering expected effects. Writing

$$Y_t^0 \equiv c_0^t(Y_0, \tilde{Z}^t, \ddot{Z}^t) \qquad Y_t^1 \equiv c_1^t(Y_0, \tilde{Z}^t, \ddot{Z}^t),$$

consider the period t expected effect

$$\Delta_{1t}(x^t) \equiv E(Y_t^1 | D^t = \delta_1^t, Y_0 = y_0, \tilde{X}^t = \tilde{x}^t) - E(Y_t^0 | D^t = \delta_1^t, Y_0 = y_0, \tilde{X}^t = \tilde{x}^t)$$

$$\equiv \tilde{\mu}_{1t}(x^t) - \tilde{\mu}_{01t}(x^t),$$

where $\tilde{X}^t \equiv (\tilde{Z}^t, W^t)$, $X^t \equiv (Y_0, \tilde{X}^t)$, $x^t \equiv (y_0, \tilde{x}^t)$ for observable proxies $\{W_t\}$. This is the analog of the period t expected effect of treatment on the treated given covariates, which here are Y_0 and $\tilde{X}^t$, but it differs from the non-dynamic concept in a significant way, as explained next.

An important and interesting feature of the present framework is that because of the dynamics involved, the effect of a natural experiment may linger beyond the end of the natural experiment ($t = T_1$). Such "lingering effects" can be substantial, especially in circumstances in which either the dynamics operate only slowly to return the dependent variable to the reference equilibrium value or the natural experiment operates to take the system to the domain of attraction of an equilibrium different from that which would have prevailed in the absence of the experiment. Consequently, to study the effects of the natural experiment it will be of interest to consider not only $t \leq T_1$, to measure "concurrent" effects, but also $t > T_1$, to measure lingering effects.

The estimation of $\tilde{\mu}_{1t}$ presents only statistical challenges, as Y_t^1 is observed given $D^t = \delta_1^t$. Estimation of $\tilde{\mu}_{01t}$ presents challenges analogous to those of the cross-section case, as Y_t^0 cannot be observed given $D^t = \delta_1^t$. Once again, however, suitable conditional independence conditions make estimation of $\tilde{\mu}_{01t}$ feasible. To see how, we write

$$\tilde{\mu}_{01t}(x^t) = \int c_0^t(y_0, \tilde{z}^t, \ddot{z}^t)\, dG_1^t(\ddot{z}^t | x^t),$$

where $G_1^t(\cdot\,|x^t) \equiv G^t(\cdot\,|\delta_1^t, x^t)$ is the distribution of $\ddot{Z}^t$ given $D^t = \delta_1^t$, $X^t = x^t$. Similarly, write $G_0^t(\cdot\,|x^t) \equiv G^t(\cdot\,|\delta_0^t, x^t)$. Given the *dynamic* CIPP condition

$$\ddot{Z}^t \perp D^t \,|\, X^t \qquad t = 1, 2, \ldots,$$

then $G_1^t = G_0^t$, and we have $\tilde{\mu}_{01t} = \tilde{\mu}_{0t}$, where

$$\tilde{\mu}_{0t}(x^t) \equiv \int c_0^t(y_0, \tilde{z}^t, \ddot{z}^t)\, dG_0^t(\ddot{z}^t \,|\, x^t).$$

By imposing suitable stationarity conditions, specifically that the conditional distribution of $\ddot{Z}^{\tau,t} \equiv (\ddot{Z}_{\tau+1}, \ldots, \ddot{Z}_{\tau+t})$ given $D^{\tau,t} \equiv (D_{\tau+1}, \ldots, D_{\tau+t})$ and $X^{\tau,t} \equiv (Y_\tau, \tilde{Z}_{\tau+1}, \ldots, \tilde{Z}_{\tau+t},\ W_{\tau+1}, \ldots, W_{\tau+t})$

does not depend on τ, together with the corresponding "full" dynamic CIPP condition

$$\ddot{Z}^{\tau,t} \perp D^{\tau,t} \mid X^{\tau,t} \qquad t = 1, 2, \ldots; \tau = 0, 1, \ldots,$$

it becomes feasible to estimate $\tilde{\mu}_{0t}$ (hence $\tilde{\mu}_{01t}$) using observations not subject to the natural experiment, that is, observations $\tau = T_1 + t, \ldots, T$.

To define an analog of unconfoundedness for the present dynamic context, write $\delta_0^{\tau,t} \equiv (\delta_{0,\tau+1}, \ldots, \delta_{0,\tau+t})$ and $\delta_1^{\tau,t} \equiv (\delta_{1,\tau+1}, \ldots, \delta_{1,\tau+t})$, let

$$\begin{aligned} Y_{\tau,t}^0 &\equiv c_0^t(Y_\tau, \tilde{Z}^{\tau,t}, \ddot{Z}^{\tau,t}) \equiv c^t(Y_\tau, \delta_0^{\tau,t}, \tilde{Z}^{\tau,t}, \ddot{Z}^{\tau,t}) \\ Y_{\tau,t}^1 &\equiv c_1^t(Y_\tau, \tilde{Z}^{\tau,t}, \ddot{Z}^{\tau,t}) \equiv c^t(Y_\tau, \delta_1^{\tau,t}, \tilde{Z}^{\tau,t}, \ddot{Z}^{\tau,t}) \end{aligned}$$

be the dynamic counterfactuals, and define $D_{\tau,t}^* = 1$ if $D^{\tau,t} = \delta_1^{\tau,t}$ and $\delta_1^{\tau,t} \neq \delta_0^{\tau,t}$ and $D_{\tau,t}^* = 0$ if $D^{\tau,t} = \delta_0^{\tau,t}$.

We define dynamic unconfoundedness for natural experiments (DUNE) as

$$(Y_{\tau,t}^0, Y_{\tau,t}^1) \perp D_{\tau,t}^* \mid X^{\tau,t} \qquad t = 1, 2, \ldots; \tau = 0, 1, \ldots.$$

The argument of Proposition 3.2 establishes that dynamic CIPP implies DUNE:

Proposition 4.1 *If* $\ddot{Z}^{\tau,t} \perp D^{\tau,t} \mid X^{\tau,t}$, $t = 1, 2, \ldots$; $\tau = 0, 1, \ldots$, *then* $(Y_{\tau,t}^0, Y_{\tau,t}^1) \perp D_{\tau,t}^* \mid X^{\tau,t}$, $t = 1, 2, \ldots$; $\tau = 0, 1, \ldots$.

Applying Rosenbaum and Rubin's (1984) results for the propensity score for each t and τ gives a dynamic analog of the propensity score,

$$p_{\tau,t}(X^{\tau,t}) \equiv P[D_{\tau,t}^* = 1 \mid X^{\tau,t}]$$

such that DUNE implies

$$(Y_{\tau,t}^0, Y_{\tau,t}^1) \perp D_{\tau,t}^* \mid p_{\tau,t}(X^{\tau,t}).$$

Unfortunately, however, in our single realization of the time series history, the sampling variation of $D_{\tau,t}^*$ provides no guide whatsoever even to $P[D_{\tau,t}^* = 1]$, much less to $p_{\tau,t}(X^{\tau,t})$. The reason is that because the natural experiments of interest here are non-recurring, the sample proportion of observations subject to the natural experiment is not a reliable estimator of the (unconditional) probability of the natural experiment. Consequently, $p_{\tau,t}$ cannot play the identical role in the present context that its analogs do in the cross-section, panel, or recurring intervention cases.

Interestingly, however, there is a modified version of the dynamic propensity score that can play a role in constructing observation weights analogous to the role played by the cross-section propensity score in the work of HIR and HI. Recall that HIR and HI used the propensity score to construct weights of the form $\hat{\omega}_t = \hat{p}(X_t)/(1-\hat{p}(X_t))$ yielding an approximation to

$$\mu_{00}(h) \equiv \int \tilde{\mu}_0(x)h(x)\, dH_0(x),$$

with h chosen as dH_1/dH_0, so that $\mu_{00}(dH_1/\, dH_0) = \mu_{01}$. In the present context, consider approximating $\tilde{\mu}_{0t}$, measuring the goodness of fit by the weighted mean squared error. Specifically, define

$$\sigma_t^2(\beta, h^t) \equiv \int (\tilde{\mu}_{0t}(x^t) - x_{\psi t}'\beta)^2 h^t(x^t)\, dH_0^t(x^t),$$

where $x_{\psi t}$ is a vector of transformations of x^t, analgous to our discussion of HI. The quantity $\sigma_t^2(\beta, h^t)$ is the weighted mean squared error of $x_{\psi t}'\beta$ as an approximation to $\tilde{\mu}_{0t}(x^t)(= \tilde{\mu}_{01t}(x^t))$ with weight function h^t, integrated over dH_0^t, the reference treatment density of X^t.

Choosing $h^t = h_{10}^t \equiv dH_1^t/\, dH_0^t$, where dH_1^t is the comparison treatment density of X^t, gives

$$\sigma_t^2(\beta, h_{10}^t) = \int (\tilde{\mu}_{0t}(x^t) - x_{\psi t}'\beta)^2 dH_1^t(x^t).$$

If β is chosen to minimize $\sigma_t^2(\beta, h_{10}^t)$, the result is a parameter vector β_{01t}^* that delivers an approximation to $\tilde{\mu}_{0t}(x^t) = \tilde{\mu}_{01t}(x^t)$ of the form $x_{\psi t}'\beta_{01t}^*$ that is optimal with respect to the comparison treatment density, dH_1^t. This implies that the prediction $X_{\psi t}'\beta_{01t}^*$ is mean-squared-error optimal as a prediction of $Y_t^0 = Y_{0,t}^0$ for the regime subject to the natural experiment, a natural optimality property in the present context.

For this weighting to be feasible theoretically, it is necessary that $dH_1^t/\, dH_0^t$ be well defined, similar to the situation in HI. (Common support suffices.) For this weighting to be empirically feasible, it is necessary that h_{10}^t can be sufficiently well estimated. A helpful condition in this regard is the stationarity condition

$$(D^{\tau,t}, X^{\tau,t}) \sim H^t, \qquad \tau = 0, 1, \ldots.$$

With this condition, dH_1^t can be estimated from observations τ subject to the natural experiment such that $D^{\tau,t} = (1,\ldots,1)$, whereas dH_0^t can be estimated from observations τ not subject to the natural experiment such that $D^{\tau,t} = (0,\ldots,0)$.

The modified propensity score mentioned above serves as a convenient device for estimating $h_{10}^t = dH_1^t/\,dH_0^t$. Suppose one has T_0 observations such that $D^{\tau,t} = (0,\ldots,0)$ and T_1 observations such that $D^{\tau,t} = (1,\ldots,1)$. Using these observations, attempt to estimate the pseudo-propensity score

$$p_t^*(X^{\tau,t}) \equiv P[D^{\tau,t} = (1,\ldots,1) \mid X^{\tau,t}]$$

using standard methods (e.g., those of HIR or HI), and denote the resulting estimate $\hat{p}_t^*$. Under suitable regularity conditions, methods analogus to those of HIR and HI can be used to show that a useful estimator of h_{10}^t is given by

$$\hat{h}_{10}^t = (T_0/T_1)\,\hat{p}_t^*/(1-\hat{p}_t^*).$$

Parallel to the approach of HI, this suggests for given t regressing $Y_{\tau+t}$ on $X_{\psi,\tau,t}$ (transforms of $X^{\tau,t}$) with weights $\hat{h}_{10}^t(X^{\tau,t})$ using only observations τ such that $D^{\tau,t} = (0,\ldots,0)$ to obtain an estimate of β_{01t}^*, say $\hat{\beta}_{01t}$. An estimate of Y_t^0 is then given by $X_{\psi t}'\hat{\beta}_{01t}$.

We refer to p_t^* as a "pseudo-propensity score" as one cannot actually estimate $P[D^{\tau,t} = (1,\ldots,1)$ $\mid X^{\tau,t}]$ from a single time-series realization. Rather, by attempting to do so using standard methods, the resulting $\hat{p}_t^*$ is useful in estimating $h_{10}^t = dH_1^t/\,dH_0^t$.

The method just outlined is an example of a "direct" method for estimating $\tilde{\mu}_{0t} = \tilde{\mu}_{01t}$. Such methods can demand a great deal from the data. The common support condition for dH_1^t and dH_0^t is not trivial, and the estimator $\hat{\beta}_{01t}$ requires sufficient observations containing histories of t observations subject to and not subject to the natural experiment. That is, T_1 and T_0 must be suitably large. Below we discuss alternative recursive or "plug-in" estimators that may be less demanding of the data.

The concepts developed in this section are now almost enough for us to state a dynamic version of Assumption B.1 suitable for studying the effects of natural experiments operating in time. To complete the necessary background, we note that for the sake of simplicity our earlier definition of the response of the dynamic settable system was more restrictive than necessary. Specifically, the responses were dependent only on their own lags, and not on the lags of other responses. Such dependence may be important in the applications of interest here. For example, it is relevant in analyzing the joint price behavior of the participants in a cartel or in understanding joint price effects in a merger.

To handle vector response dynamics, let $\mathcal{J}_1, \mathcal{J}_2, \ldots$ be a countable partition of $\mathbb{N}$ such that for all j in $\mathcal{J}_h$ the responses $\mathcal{X}_j(0, \cdot)$, are determined as $\mathcal{X}_j(0, \cdot) = \{Y_{jt}\}$, generated as

$$\mathbf{Y}_{ht} = \mathbf{c}_{ht}(\mathbf{Y}_h^{t-1}, Z_{[h]}^t), \qquad t = 1, 2, \ldots,$$

where $\mathbf{Y}_{ht}$ is the vector with elements Y_{jt}, $j \in \mathcal{J}_h$, $\mathbf{c}_{ht}$ is a vector-valued function, $\mathbf{Y}_h^{t-1} \equiv (\mathbf{Y}_{h0}, \mathbf{Y}_{h1}, \ldots, \mathbf{Y}_{ht-1})$, and $Z_{[h]}^t = (Z_{[h],1}, \ldots, Z_{[h],t})$, where $Z_{[h],t}$ is the countable vector containing Z_{it}, $i \notin \mathcal{J}_h$. Substitutions identical to those at the outset of this section yield a cumulative response function for $j \in \mathcal{J}_h$ of the form

$$Y_{jt} = c_j^t(\mathbf{Y}_{h,0}, Z_{[h]}^t).$$

We can now state the following dynamic analog of Assumption B.1:

Assumption C.1 (Data Generating Process) *Let* $(\mathcal{Y}, \mathcal{D}, \mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ *be a dynamic settable system such that* (i) $\mathcal{Y}$ *is a vector-valued sequence with initial element* $\mathcal{Y}_0$, $\mathcal{D}$ *is a* $\{0,1\}$*-valued sequence,* $\mathcal{W}$ *is an* $\mathbb{R}^\ell$*-valued sequence,* $\ell \in \mathbb{N}$, $\tilde{\mathcal{Z}}$ *is an* $\mathbb{R}^k$*-valued sequence,* $k \in \mathbb{N}$, *and* $\ddot{\mathcal{Z}}$ *is an* $\mathbb{R}^\infty$*-valued sequence*; (ii) $\mathcal{Y} \not\Rightarrow (\mathcal{D}, \mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, $\mathcal{D} \not\Rightarrow (\mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$, *and* $\mathcal{W} \not\Rightarrow \mathcal{Y}$; (iii) $(\mathcal{Y}_0, \mathcal{D}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ *generates random variables* $(Y_0, \{(D_t, \tilde{Z}_t, \ddot{Z})\})$ *that determine the vector-valued responses* $\{Y_t\}$ *of* $\mathcal{Y}$ *according to*

$$Y_t = c^t(Y_0, D^t, \tilde{Z}^t, \ddot{Z}^t), \qquad t = 1, 2, \ldots,$$

where $\{c^t\}$ *is a sequence of unknown vector-valued functions*; (iv) *for* $t = 1, 2, \ldots$, *and* $\tau = 0, 1, \ldots$,*the joint distribution of* $(D^{\tau,t}, X^{\tau,t})$ *is* H^t, *the conditional distribution of* $\ddot{Z}^{\tau,t}$ *given* $(D^{\tau,t}, X^{\tau,t}) = (d^t, x^t)$ *is* $G^t(\cdot \mid d^t, x^t)$; *and* $\ddot{Z}^{\tau,t} \perp D^{\tau,t} \mid X^{\tau,t}$; (v) *The realizations of* Y_t, D_t, *and* X_t *are observed, whereas those of* $\ddot{Z}_t$ *are not.*

Assumption C.1 defines a data generating process admitting formal causal structure and vector response dynamics. By suitably restricting the dependence of $\{(D_t, W_t, \tilde{Z}_t, \ddot{Z}_t)\}$ by appropriate choice of P, and by suitably controlling the dependence of $\{Y_t\}$ by suitable restrictions on the functions $\{c^t\}$, it is possible to state regularity conditions (e.g., suitable and considerably generalized analogs of A.2) that will support parametric or nonparametric estimation of $\tilde{\mu}_{1t}$ and $\tilde{\mu}_{01t}$. Specifically, it is possible to provide regularity conditions of this sort ensuring consistency, convergence rate, asymptotic distribution, and/or consistency of asymptotic variance estimator results, etc., for a range of estimators $\hat{\mu}_{1t}$ and $\hat{\mu}_{01t}$ for $\tilde{\mu}_{1t}$ and $\tilde{\mu}_{01t}$ respectively. Conditions for the parametric case ensuring consistency, asymptotic normality, and consistent covariance estimation are available from work of Gallant and White (1988), for example. For conditions for the nonparametric case, see Chen (2005) and Chen and White (1999).

Note that the stationarity and full dynamic CIPP conditions imposed in (iv) facilitate estimation of $\tilde{\mu}_{01t} = \tilde{\mu}_{0t}$ and of dH_1^t / dH_0^t, as discussed above.

Given asymptotic results for $\hat{\mu}_{1t}$ and $\hat{\mu}_{01t}$ jointly, one can then derive asymptotic results for estimators of the period t expected effect of the treatment $\Delta_{1t} = \tilde{\mu}_{1t} - \tilde{\mu}_{01t}$ having the form

$$\hat{\Delta}_{1t} = \hat{\mu}_{1t} - \hat{\mu}_{01t},$$

as well as for average effect estimators such as

$$\bar{\Delta}_1^* \equiv T^{*^{-1}} \sum_{t \in \mathcal{T}^*} \hat{\Delta}_{1t}(X^t) = \bar{\mu}_1^* - \bar{\mu}_{10}^*,$$

where $\mathcal{T}^*$ is a given set of time indexes having T^* elements, and

$$\bar{\mu}_1^* \equiv T^{*^{-1}} \sum_{t \in \mathcal{T}^*} \hat{\mu}_{1t}(X^t) \qquad \bar{\mu}_{01}^* = T^{*^{-1}} \sum_{t \in \mathcal{T}^*} \hat{\mu}_{01t}(X^t).$$

In particular cases, a suitable estimator $\bar{\mu}_1^*$ may be simply $\bar{\mu}_1^* = T^{*^{-1}} \sum_{t \in \mathcal{T}^*} Y_t$. Analogous discounted sums may also be of interest and can be estimated in the obvious way. The space available here does not permit a detailed analysis of any particular estimator in the dynamic case. Instead, we consider certain broadly relevant aspects of estimation strategy.

Given dynamic CIPP, the estimation problem reduces to estimating the conditional expectations

$$\tilde{\mu}_{0t}(X^t) = E(Y_t^0 | D^t = \delta_0^t, X^t), \quad t = 1, 2, \ldots,$$

as dynamic CIPP ensures $\tilde{\mu}_{0t} = \tilde{\mu}_{01t}$. (The considerations for $\tilde{\mu}_{1t}$ are analogous, so we do not provide explicit discussion.) Estimating these objects can be viewed as a version of the problem of creating "multi-step" or "multiple horizon" predictions, as we now discuss.

Multiple horizon prediction is a time-series topic of long-standing interest. (See, for example, recent work of Ing (2002, 2004) and articles cited therein.) The area is sufficiently developed that most forecasting software now has standard commands for producing multi-step forecasts, although not all produce estimates of the version relevant here. Moreover, in particular applications it may be desirable to implement methods more sophisticated to varying degrees than standard methods. The structure of the problem typically makes the development and implementation of such methods straightforward.

Suitable analogs of the cross-section methods of Hahn, HIR, HI, LRW, and of Heckman and his collaborators in fact provide a rich body of technique that can serve as a starting point for developing a wide array of methods, parametric and nonparametric, specifically tailored to estimating the conditional expectations relevant to assessing the effects of natural experiments in time-series contexts. The opportunities are sufficiently broad that here we touch only on some of the central issues.

Above we noted that estimation of $\tilde{\mu}_{0t}$ involves versions of the multi-step prediction problem. In the standard problem, forecasts are made for multiple horizons in the future ($t = 1, 2, \ldots$) based on information available when the forecast is made ($t = 0$). In the present application, histories of the "future" covariates ($\tilde{Z}^t, W^t, t = 1, 2, \ldots$) are also available to inform the prediction. Although this distinction can have major implications for the behavior of the resulting forecasts, the conceptual distinction is modest, and the insights and approaches of the multi-step prediction literature apply either directly or with straightforward modification.

Broadly speaking, there are two methods for creating multiple horizon forecasts: the direct approach (an example of which we considered above) and the recursive or "plug-in" approach. The direct approach estimates a separate prediction model for each horizon t. The recursive approach builds up a prediction period by period, typically using some assumed Markov structure. In general, these approaches may be purely parametric, purely nonparametric, or may blend parametric and nonparametric elements. The recursive approach may require evaluation of certain integrals, which may be done analytically, by numerical approximation, or by simulation.

The direct approach is straightforward, but can be somewhat labor-intensive, given that the modeling process must begin anew for each forecast horizon, t. As we saw in our earlier example, it is in principle amenable to methods exploiting the pseudo-propensity score. The resulting forecasts may be increasingly unreliable at greater forecast horizons, however, as fewer observations are available for estimation of longer horizon relationships. Ultimately, the number of steps ahead for which predictions can be produced by the direct approach is limited by this feature of the process.

The recursive approach is also straightforward and much less labor intensive, as it typically requires estimation of only one-step-ahead prediction relationships. This makes it less demanding of the data for estimation purposes, and it can generate predictions at any horizon for which covariates are available. On the other hand, the recursive approach may require the validity of certain simplifying assumptions whose failure might affect the resulting predictions. Nor are the possibilities for exploiting the pseudo-propensity score as straightforward in the recursive case as they are in the direct case.

As we have already discussed an example of the direct approach to estimating $\tilde{\mu}_{0t}$, we now consider recursive estimation in greater detail. To recursively estimate $\tilde{\mu}_{0t}$, one can make use of the fact that

$$E(Y_t^0|X^t = x^t) = \int y_t dF_{0t}(y_t|x^t),$$

where $dF_{0t}(\cdot|x^t)$ is the conditional density of Y_t^0 given $X^t = x^t$. The properties of conditional densities give an analog of the discrete time forward Chapman-Kolmogorov equations

$$\begin{aligned} dF_{0t}(y_t|x^t) &= \int dF_{0t|t-1}(y_t|y_{t-1}, x^t)\, dF_{0t-1}(y_{t-1}|x^t) \\ &= \int dF_{0t|t-1}(y_t|y_{t-1}, x^t)\, dF_{0t-1}(y_{t-1}|x^{t-1}), \end{aligned}$$

whenever the conditional independence condition $dF_{0t-1}(y_{t-1}|x^t) = dF_{0t-1}(y_{t-1}|x^{t-1})$ holds, where $dF_{0t|t-1}(\cdot|y_{t-1}, x^t)$ is the conditional density of Y_t^0 given $Y_{t-1}^0 = y_{t-1}$, $X^t = x^t$. Imposing a Markov condition of the form

$$dF_{0t|t-1}(y_t|y_{t-1}, x^t) = dF^s_{0t|t-1}(y_t|y_{t-1}, \tilde{x}^t_{t-s}),$$

where $\tilde{x}^t_{t-s} \equiv (\tilde{x}_{t-s}, \ldots, \tilde{x}_t)$, $\tilde{x}_t \equiv (\tilde{z}_t, w_t)$, one can recursively generate an estimate of dF_{0t} and therefore $\tilde{\mu}_{0t}$ using an estimate of $dF^s_{0t|t-1}$, say $d\hat{F}^s_{0t|t-1}$, based on data not subject to the natural experiment. The recursion is then

$$\begin{aligned} d\hat{F}_{0t}(y_t|x^t) &= \int d\hat{F}^s_{0t|t-1}(y_t|y_{t-1}, \tilde{x}^t_{t-s})\, d\hat{F}_{0t-1}(y_{t-1}|x^{t-1}) \\ \hat{\mu}_{0t}(x^t) &= \int y_t d\hat{F}_{0t}(y_t|x^t), \quad t = 1, 2, \ldots. \end{aligned}$$

The approach just described may be implemented parametrically or nonparametrically. Specific parametric assumptions can dramatically simplify the required computations, permitting direct recursive calculation of the estimates. For example, if

$$E(Y_t^{0\prime}|Y_{t-1}^0, Y_{t-2}^0, \ldots; X^t) = Y_{t-1}^{0\prime}\rho_0 + X_{\psi t}^{\prime}\beta_0,$$

for unknown coefficient matrices ρ_0 and β_0, then an estimate of $E(Y_t^{0\prime}|X^t)$ can be constructed recursively as

$$\hat{Y}_t^{0\prime} = \hat{Y}_{t-1}^{0\prime}\hat{\rho}_0 + X_{\psi t}^{\prime}\hat{\beta}_0 \quad t = 1, 2, \ldots,$$

where $\hat{\rho}_0$ and $\hat{\beta}_0$ are estimated from data not subject to the natural experiment and $\hat{Y}_0^0$ is the initial value, Y_0. That is, the predictions are "rolled forward" and updated with current covariate information.

An alternative to the parametric approach is simulation of predictive distributions. Work of Swanson and Urbach (2005), Korenok and Swanson (2005), and Corradi, Distaso, and Swanson (2005) contains methods of this sort that with suitable adaptations may be highly effective.

The discussion here has barely scratched the surface of the possibilities. We have not gone into the specifics of any of the suggested estimators, nor have we paid explicit attention to the variety of relevant ways that time-series data may be generated. For example, the covariates may be integrated, and if so, the responses Y_t may be cointegrated with the covariates. The framework provided by Assumption D.1 indeed applies to such situations, and the methods discussed here apply, to the extent that the objects of interest (e.g. the $\tilde{\mu}_{0t}$'s) exist and can be consistently estimated.

Concerning this last point, an important caveat is that successful estimation of $\tilde{\mu}_{01t}(X^t)$ using $\hat{\mu}_{0t}(X^t)$ requires suitable behavior for the support of X^t in the treatment (natural experiment) and non-treatment regimes. As stressed by Heckman, Ichimura, and Todd (1998), reliable estimates obtain when the support for the treated regime coincides with or is a subset of that for the non-treated regime, and not necessarily otherwise.

For concreteness and brevity, we have focused attention in this section almost entirely on $\Delta_{1t}(X^t)$ and its components. Nevertheless, other expected effects may well be of interest. For example, in retrospective analyses, the entire available history of covariates and not just the history up to time t may be relevant in assessing expected effects. A version of the smoothed Kalman filter may be useful in such cases. As stated earlier, it is feasible here merely to broadly map the territory and thereby indicate promising directions for future research. Accordingly, we leave the interesting possibilities mentioned above and related investigations for later exploration.

5 Justifying Predictive Proxies and Testing CIPP

Given the key role played by CIPP or dynamic CIPP in ensuring unconfoundedness, it is crucial to critically examine its plausibility in each application. The situation is closely parallel to that of instrumental variables (IV) estimation, where the validity of the chosen instrumental variables must be critically examined. Parallel to the IV case, in which the validity of the identifying instruments cannot be tested but can only be theoretically justified, here it is also true that the validity of the predictive proxies cannot be tested, but can only be theoretically justified. In the IV case, however, the validity of over-identifying instrumental variables can be tested, and here too, a parallel exists: it is possible to test CIPP. In this section we discuss validating predictive proxies and testing CIPP. In particular, we discuss how knowledge of the underlying economic structure can assist in justifying unconfoundedness. For simplicity, the discussion focuses primarily on the non-dynamic case, but we give formal results applying to the full dynamic case.

5.1 Validating Predictive Proxies and Assessing Properness

Recall that the auxiliary variables $\mathcal{W}$ are predictive proxies if they are statistically related to $\ddot{\mathcal{Z}}$ and do not cause $\mathcal{Y}$. They are valid predictive proxies if neither $\mathcal{Y}$ nor $\mathcal{D}$ cause $\mathcal{W}$. We refer to these requirements as the "validity conditions" for predictive proxies.

Although it is not possible to affirmatively prove that a given $\mathcal{W}$ is valid, it is possible to "justify" $\mathcal{W}$ by checking the validity conditions for conformity with economic theory and examining the plausibility of properness for the resulting covariates. Elements of $\mathcal{W}$ plausibly failing any of the validity conditions can be viewed as invalid and removed from further consideration. The identification of important unobserved causes that have not been adequately proxied is the main theoretical tool for calling into question the properness of a given set of covariates.

The role played by economic theory in assessing the causal requirements for validity of $\mathcal{W}$ is usually immediately apparent in applications, as causality relations reduce to functional dependence in the present context. Interestingly, the statistical dependence requirement can also be cast as a causality requirement, making it equally susceptible to the application of domain knowledge and economic theory in assessing its plausibility.

According to Reichenbach's (1956) *common causality principle*, there are three channels for statistical dependence to arise between $\mathcal{W}$ and $\ddot{\mathcal{Z}}$: (1) $\mathcal{W}$ can cause $\ddot{\mathcal{Z}}$; (2) $\ddot{\mathcal{Z}}$ can cause $\mathcal{W}$; (3) $\ddot{\mathcal{Z}}$ and $\mathcal{W}$ can have a common cause, say $\mathcal{S}$. In our settable system framework, for channel (1) the dependence is between $\mathcal{W}(1,\cdot)$ and $\ddot{\mathcal{Z}}(0,\cdot)$; for channel (2), the dependence is between $\mathcal{W}(0,\cdot)$ and $\ddot{\mathcal{Z}}(1,\cdot)$; and for channel (3) the dependence is between $\mathcal{W}(0,\cdot)$ and $\ddot{\mathcal{Z}}(0,\cdot)$, driven by underlying settings $\mathcal{S}(1,\cdot)$. There is apparently also a fourth channel for dependence in our framework, namely dependence between settings $\mathcal{W}(1,\cdot)$ and $\ddot{\mathcal{Z}}(1,\cdot)$ governed by P. If $\mathcal{W}(1,\cdot)$ and $\ddot{\mathcal{Z}}(1,\cdot)$ are truly set according to P, however, then the randomly selected element ω of Ω determining the realizations $(\mathcal{W}(1,\omega), \ddot{\mathcal{Z}}(1,\omega))$ can be viewed as a common cause, which reduces the apparent fourth channel to the third channel.

Accordingly, it suffices to examine Reichenbach's three channels for appropriate sources of statistical dependence between $\mathcal{W}$ and $\ddot{\mathcal{Z}}$. Now if $\mathcal{W}$ causes $\ddot{\mathcal{Z}}$, one can substitute $\mathcal{W}$ and its sufficient concomitants for $\ddot{\mathcal{Z}}$ into the response function, yielding a modified response function containing $\mathcal{W}$ as observed causes of $\mathcal{Y}$, so that $\mathcal{W}$ is properly a sub-vector of a modified version of $\tilde{\mathcal{Z}}$. Channel (1) cannot therefore be the source of the required statistical dependence, as this violates $\mathcal{W} \not\rightsquigarrow \mathcal{Y}$. On the other hand, channel (2) ($\ddot{\mathcal{Z}} \rightsquigarrow \mathcal{W}$) offers no conflict with the other requirements for validity; it is thus a legitimate source of the statistical dependence requirement. Finally, consider channel (3), and assume without loss of generality that the common cause $\mathcal{S}$ is unobserved. If $\mathcal{S}$ causes $\ddot{\mathcal{Z}}$, then one can substitute $\mathcal{S}$ and its sufficient concomitants for $\ddot{\mathcal{Z}}$ into the response function, yielding a modified response function in which the modified version of $\ddot{\mathcal{Z}}$ contains $\mathcal{S}$ as a sub-vector. As $\mathcal{S}$ also causes $\mathcal{W}$, we again have statistical dependence arising from the operation of channel (2). It follows that the settable system can be standardized (by the substitutions just described) to permit the validity conditions for predictive proxies to be solely causal: (i) $\ddot{\mathcal{Z}} \rightsquigarrow \mathcal{W}$; (ii) $\mathcal{W} \not\rightsquigarrow \mathcal{Y}$; (iii) $\mathcal{Y} \not\rightsquigarrow W$; (iv) $\mathcal{D} \not\rightsquigarrow \mathcal{W}$. Valid covariates are thus either valid predictive proxies $\mathcal{W}$ or valid observable causes $\tilde{\mathcal{Z}}$ of $\mathcal{Y}$, that is, $\tilde{\mathcal{Z}}$ satisfies: (i) $\tilde{\mathcal{Z}} \rightsquigarrow Y$; (ii) $\mathcal{Y} \not\rightsquigarrow \tilde{\mathcal{Z}}$; (iii) $\mathcal{D} \not\rightsquigarrow \tilde{\mathcal{Z}}$.

We discuss construction of valid predictive proxies in further detail in the next sub-section. For now it suffices to recognize that a leading possibility for the construction of valid predictive proxies is when $\mathcal{W}$ is determined as an error-laden measurement of the setting of $\ddot{\mathcal{Z}}$.

An example will help provide insight into the process of validating predictive proxies and assessing the plausibility of CIPP. We start from a situation in which the natural experiment is the only observed cause, so that

$$Y_t = c(D_t, \ddot{Z}_t).$$

At this stage there are no predictive proxies.

Now consider whether to augment the predictive proxies. We must do so if there is any evidence that the distribution of $\ddot{Z}_t$ differs between the two regimes as CIPP fails in the presence of such a shift. (Here CIPP is equivalent to independence of $\ddot{Z}_t$ and D_t.) Although $\ddot{Z}_t$ is unobserved, indirect evidence of distributional shifts is often available from observable proxies for $\ddot{Z}_t$. In Section 5.2, we discuss tests based on this fact. For now, we proceed assuming the possibility of such a shift.

Taking the cartel example for concreteness, suppose the unobserved average price of an important raw material is higher in the cartel regime for reasons unrelated to the cartel. Then we will observe an increase in average price of the cartelized product, but in the absence of any predictive proxy, there is no way to separate the effect of the cartel from the unrelated effect of the cost increase. That is, CIPP fails as a result of the unaccounted-for change in the raw material price distribution.

Although the actual raw material price may not be available (for example, because accurate records are not cost effective for the firm to keep), the economist may have access to a monthly price index for the commodity. These data are not the true prices driving firm decisions, but they are an appealing candidate proxy. Just as the true price must be unaffected by the cartel, so must its proxy be unaffected. This is often plausible, as price indexes are typically weighted averages of prices surveyed in some standard way. As long as both the underlying surveyed prices and the weights used are not appreciably impacted by the cartel, then we are justified in adding the price index to the set of valid predictive proxies, now designated W_t^1.

The next step is to again ask whether components of $\ddot{Z}_t$ have distributions (conditional now on W_t^1) that differ between regimes. If not, we are done. If so, CIPP is in question, and further steps are required.

Continuing the example, suppose there is a possible shift in demand between regimes unrelated to the cartel and not adequately proxied by the price index, W_t^1. For example, suppose the cartel's product is widely used by a broad array of "downstream buyers," and that an increase in demand for downstream buyers' products causes an increase in demand for the cartel's product.

To belong to $\ddot{Z}_t$, any demand shifters must be unaffected by the cartel. This is plausible if the cartelized industry is of relatively modest size. As it is typically difficult to measure the ideal demand shifters, we must seek predictive proxies. Thus, consider using as predictive proxies indexes of industrial production (IP) for industry segments representing the downstream buyers.

If the IP index is for a sector broad enough to contain the cartel, the IP index is not a valid predictive proxy, as it is driven by the cartel activity. If the cartel is only a small part of the covered sector then the problem is mitigated. Better, however, is to remove the cartel component from the index.

Alternatively, if the IP index is for a sector for which a quantity restriction by the cartel would curtail production, then the index for that sector is again driven by the cartel activity and validity is in question. In this case one can seek proxies for demand for the downstream buyers' products.

Given valid additional predictive proxies, they may be appended to W_t^1 to create W_t^2, and the process iterated, until one arrives finally at some collection of covariates $X_t \equiv (\tilde{Z}_t, W_t)$ that have been justified to the extent that (a) the predictive proxies are plausibly valid; and (b) CIPP plausibly holds.

As can be shown (see White, 2005a, Proposition 5.2), the impact of the violation of CIPP is a matter of degree. The more important the unobserved cause and the greater is $G_1 - G_0$, the greater is the adverse impact on the estimation of the effect of interest. Thus, the greatest care must be directed toward obtaining valid proxies for the most important unobserved causes and toward those for which $G_1 - G_0$ is potentially greatest. Neglect of minor causes or those for which $G_1 - G_0$ is small can be shown to have minor impact.

Variables invalid as predictive proxies have no place in the analysis. This applies particularly to variables not satisfying validity condition (i), that is, variables not having well-justified economic links to $\ddot{Z}_t$. Economic time series are sufficiently numerous that one can often find a series with just the right pattern to absorb the effect of the natural experiment. An apparent statistical dependence may exist between the supposed proxy and $\ddot{\mathcal{Z}}$, but this may be more apparent than real, an artifact of the data mining involved in finding such a series. A strong antidote to such data mining is economic theory. The absence of a viable economic justification for including a particular proxy and, in particular, a plausible causal link from $\ddot{\mathcal{Z}}$ to $\mathcal{W}$, is strong grounds for its exclusion, especially given how easily spurious proxies can mask the true effect.

Economic theory thus plays a decisive role in isolating and analyzing the effects of a natural experiment. Specifically, economic theory identifies: (1) the causal factors $(\tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ not caused by the dependent variable or the natural experiment; (2) variables determined by the dependent variable or natural experiment and that are therefore *not* legitimate components of $\tilde{\mathcal{Z}}$ or $\ddot{\mathcal{Z}}$; (3) variables $\mathcal{W}$ that are valid predictive proxies for $\ddot{\mathcal{Z}}$; and (4) variables that are *not* valid proxies for $\ddot{\mathcal{Z}}$.

5.2 Testing CIPP

Although CIPP cannot be affirmatively verified, it can be empirically tested for any given covariates, using additional valid predictive proxies.

We begin with covariates $X_t \equiv (\tilde{Z}_t, W_t)$ used to construct $X_{\psi t}$ as above. (Note that here ψ may be different, but we maintain notation for simplicity.) We explicitly allow $X_{\psi t}$ to contain only the constant.

Suppose we have additional proxies V_t for $\ddot{Z}_t$, not included in X_t. For example, in the cartel example, we may have X_t containing certain raw material prices and demand shifters, but we may have further proxies V_t for other raw material prices and demand shifters not included in X_t.

To isolate any failure of CIPP, we require certain identifying information. The following structure suffices: we suppose that additional proxies $\mathcal{V}$ have responses generated by W_t, $\tilde{Z}_t$, $\ddot{Z}_t$, and further unobservables U_t as

$$V_t = v(\ddot{Z}_t, U_t, \tilde{Z}_t, W_t), \qquad t = 1, 2, \ldots,$$

for some unknown response function v. If V_t is determined solely by $\ddot{Z}_t$ and U_t, we can view V_t as a version of $\ddot{Z}_t$ contaminated by measurement errors U_t.

The properties required of the settable variable $\mathcal{U}$ yielding U_t are: (i) $\mathcal{Y} \not\rightsquigarrow \mathcal{U}$; (ii) $\mathcal{U} \not\rightsquigarrow \mathcal{Y}$; and (iii) $\mathcal{D} \not\rightsquigarrow \mathcal{U}$. These are reasonable conditions for measurement errors; they also help ensure that $\mathcal{V}$ is a valid predictive proxy.

With one further condition we have a means of identifying and testing failures of CIPP. This is the conditional independence condition

$$U_t \perp D_t \mid \ddot{Z}_t, \tilde{Z}_t, W_t,$$

a plausible condition for measurement errors. Let $J(\cdot|\ddot{z}, x)$ denote the distribution for U_t given $(\ddot{Z}_t, X_t) = (\ddot{z}, x)$. The conditional expectation of V_t given $X_t = x$ in regime 0 is then

$$\int v(\ddot{z}, u, x)\, dJ(u \mid \ddot{z}, x)\, dG_0(\ddot{z} \mid x)$$

and in regime 1 is

$$\int v(\ddot{z}, u, x)\, dJ(u \mid \ddot{z}, x)\, dG_1(\ddot{z} \mid x).$$

If CIPP holds ($G_0 = G_1$), then these conditional expectations are identical. If we find empirical evidence that equality fails, we have evidence that CIPP fails.

Observe the roles played by our assumptions on v, $\mathcal{U}$, and J. If v depended on D_t, if $\mathcal{U}$ were caused by $\mathcal{D}$, or if J were not stable across regimes, the conditional expectation of V_t given X_t would differ across regimes and the failure of CIPP would not be the sole and necessary reason for this instability.

The stability of this conditional expectation is a necessary and not a sufficient condition for CIPP. Thus, tests based on this stability could fail to detect violations of CIPP. To gain additional power one can test further necessary conditions. In particular, let ϕ be any measurable $r \times 1$ vector-valued function of V_t. Then $E(\phi(V_t)|X_t)$ is also stable across regimes given CIPP; the common conditional expectation is

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

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

To construct a test, we exploit the flexible approximation capabilities of $X_{\psi t}$: suppose there exist matrices $a^*_{\phi,0}$ and $a^*_{\phi,1}$ such that

$$\begin{aligned}
\int \phi(v(\ddot{z}, u, x))' dJ(u \,|\, \ddot{z}, x)\, dG_0(\ddot{z}|x) &= x'_\psi a^*_{\phi,0} \\
\int \phi(v(\ddot{z}, u, x))' dJ(u \,|\, \ddot{z}, x)\, dG_1(\ddot{z} \,|\, x) &= x'_\psi a^*_{\phi,1}.
\end{aligned}$$

Under the CIPP null hypothesis

$$H_0 : G_0 = G_1 \qquad a.s.$$

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,

$$\begin{aligned}
\hat{\alpha}_{\phi,0} &\equiv (\sum_{t=1}^{T} (1 - D_t) X_{\psi t} X'_{\psi t})^{-1} \sum_{t=1}^{T} (1 - D_t) X_{\psi t}\, \phi(V_t)' \\
\hat{\alpha}_{\phi,1} &\equiv (\sum_{t=1}^{T} D_t X_{\psi t} X_{\psi t})^{-1} \sum_{t=1}^{T} D_t X_{\psi t}\, \phi(V_t)'.
\end{aligned}$$

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

$$\mathcal{W}_T \equiv T\ [\text{vec}\ (\hat{\alpha}_{\phi,0} - \hat{\alpha}_{\phi,1})]' \hat{C}^{-1}\ [\text{vec}\ (\hat{\alpha}_{\phi,0} - \hat{\alpha}_{\phi,1})],$$

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

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

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

The test statistic can be computed by exploiting the regression representation

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

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

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

or

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

where $\tilde{Z}_{\psi t} \equiv (I \otimes X_{\psi t})$, $\tilde{a}^*_{\phi,0} \equiv$ vec $a^*_{\phi,0}$, $\tilde{a}^*_{\phi,1} \equiv$ vec $a^*_{\phi,1}$. To test CIPP it thus suffices to regress $\phi(V_t)$ on $Z_{\psi t}$ and $D_t Z_{\psi t}$ and test whether the coefficients on $D_t Z_{\psi t}$ (that is, $\tilde{a}^*_{\phi,1} - \tilde{a}^*_{\phi,0}$) are jointly zero. This is straightforward, but care should be taken to use a suitable heteroskedasticity-and-autocorrelation consistent (HAC) covariance estimator in constructing the test statistic (e.g. Gonçalves and White, 2005).

To test dynamic CIPP, we extend the function v to admit lags as follows:

$$V_t = v(\ddot{Z}^t_{t-s}, U_t, \tilde{Z}^t_{t-s}, W^t_{t-s})$$

for a given $s \in \mathbb{N}$, where $\ddot{Z}^t_{t-s} \equiv (\ddot{Z}_{t-s}, \ldots, \ddot{Z}_t)$ and similarly for $\tilde{Z}^t_{t-s}$ and W^t_{t-s}. Thus, a finite number of lags of $\ddot{Z}_t$, $\tilde{Z}_t$, and W_t explicitly determine V_t, together with U_t . As U_t is a countably-dimensioned vector, it may implicitly contain lags. The presence of only a finite number of lags is a convenience to render v time-invariant, necessitated by operating in a framework in which $t = 1, 2, \ldots$. Consideration of a doubly infinite index $t = 0, \pm 1, \pm 2, \ldots$ would permit infinite histories of all arguments.

We now extend Assumption C.1 to accommodate sequences $\mathcal{V}$.

Assumption D.1 (Data Generating Process) *Let* $(\mathcal{Y}, \mathcal{D}, \mathcal{U}, \mathcal{V}, \mathcal{W}, \tilde{\mathcal{Z}}, \ddot{\mathcal{Z}})$ *be a dynamic settable system such that:* (i) $\mathcal{Y}$ *is a vector-valued sequence with initial element* $\mathcal{Y}_0$, $\mathcal{D}$ *is a* $\{0,1\}$*-valued sequence,* $\mathcal{U}$ *is an* $\mathbb{R}^\infty$*-valued sequence,* $\mathcal{V}$ *is an* $\mathbb{R}^m$*-valued-sequence,* $m \in \mathbb{N}$, $\mathcal{W}$ *is an* $\mathbb{R}^\ell$*-valued sequence,* $\ell \in \mathbb{N}$, $\tilde{\mathcal{Z}}$ *is an* $\mathbb{R}^k$*-valued sequence,* $k \in \mathbb{N}$, *and* $\ddot{\mathcal{Z}}$ *is an* $\mathbb{R}^\infty$*-valued sequence;* (ii) $Y \rightsquigarrow\!|$

$(\mathcal{D},\mathcal{U},\mathcal{V},\tilde{\mathcal{W}},\tilde{\mathcal{Z}},\ddot{\mathcal{Z}})$, $\mathcal{D} \not\Rightarrow (\mathcal{U},\mathcal{V},\mathcal{W},\tilde{\mathcal{Z}},\ddot{\mathcal{Z}})$, and $(\mathcal{U},\mathcal{V},\mathcal{W}) \not\Rightarrow \mathcal{Y}$; (iii) *$(\mathcal{Y}_0,\mathcal{D},\tilde{\mathcal{Z}},\ddot{\mathcal{Z}})$ generates random variables $(Y_0,\{(D_t,\tilde{Z}_t,\ddot{Z}_t)\})$ that determine the responses $\{Y_t\}$ of $\mathcal{Y}$ according to*

$$Y_t = c^t(Y_0, D^t, \tilde{Z}^t, \ddot{Z}^t), \qquad t = 1,2,\ldots,$$

where $\{c^t\}$ is a sequence of unknown measurable vector-valued functions; and $(\ddot{\mathcal{Z}},\mathcal{U},\tilde{\mathcal{Z}},\mathcal{W})$ generates random variables $(\ddot{Z}_t, U_t, \tilde{Z}_t, W_t)$ that determine the responses V_t of $\mathcal{V}$ according to

$$V_t = v(\ddot{Z}^t_{t-s}, U_t, \tilde{Z}^t_{t-s}, W^t_{t-s}), \qquad t = s+1, s+2, \ldots,$$

where v is an unknown $\mathbb{R}^m$-valued measurable function for some $\tau \geq 0$; (iv) *for $t = 1,2,\ldots$, $\tau = 0,1,\ldots$, the joint distribution of $(D^{\tau,t}, X^{\tau,t})$ is H^t; the conditional distribution of $\ddot{Z}^{\tau,t}$ given $(D^{\tau,t}, X^{\tau,t}) = (d^t, x^t)$ is $G^t(\cdot|d^t,x^t)$; and, defining $\tilde{X}^t_{t-s} \equiv (\tilde{Z}^t_{t-s}, W^t_{t-s})$, the conditional distribution of $U_{\tau+t}$ given $D^{\tau+t}_{\tau+1} = d^t$, $\ddot{Z}^{\tau+t}_{\tau+t-s} = \ddot{z}^t_{t-s}$, $\tilde{X}^{\tau+t}_{\tau+t-s} = \tilde{x}^t_{t-s}$ is $J(\cdot|\ddot{z}^t_{t-s}, \tilde{x}^t_{t-s})$;* (v) *The realizations of Y_t, D_t, V_t, W_t, and $\tilde{Z}_t$ are observed, whereas those of U_t and $\ddot{Z}_t$ are not.*

For the next assumption, we define

$$\hat{M}_0 \equiv T_0^{-1} \sum_{t \in \mathcal{T}_0} X_{\psi t} X'_{\psi t} \qquad \hat{M}_1 \equiv T_1^{-1} \sum_{t \in \mathcal{T}_1} X_{\psi t} X'_{\psi t},$$

where $\mathcal{T}_0$ and $\mathcal{T}_1$ are suitable disjoint subsets of $\{1,\ldots,T\}$ such that $\mathcal{T}_0$ contains reference treatment data ($D_t = \delta_{0t}$) and $\mathcal{T}_1$ contains natural experiment data ($D_t = \delta_{1t} \neq \delta_{0t}$). $\mathcal{T}_0$ contains T_0 observations and $\mathcal{T}_1$ contain T_1 observations, $T_0 + T_1 \leq T$. The regressors $X_{\psi t}$ now contain transformations of $\tilde{X}^t_{t-s}$. We also define $\hat{p}_0 = T_0/T$, $\hat{p}_1 \equiv T_1/T$, and

$$\hat{K}_{\phi,0} \equiv T_0^{-1} \sum_{t \in \mathcal{T}_0} X_{\psi t} \phi(V_t)' \qquad \hat{K}_{\phi,1} \equiv T_1^{-1} \sum_{t \in \mathcal{T}_1} X_{\psi t} \phi(V_t)'.$$

We denote the joint distribution of $\ddot{Z}^t_{t-s}$ given $t \in \mathcal{T}_i$ and $\tilde{X}^t_{t-s} = \tilde{x}^t_{t-s}$ by $G^t_{i,t-s}(\cdot|\tilde{x}^t_{t-s})$, $i = 0,1$.

We specify law of large numbers and central limit requirements as follows.

Assumption D.2 *For given $q \in \mathbb{N}$ and given known measurable scalar-valued functions $\psi_0 = 1$, ψ_j, $j = 1,\ldots,q-1$, let $X_{\psi t} \equiv (\psi_0(\tilde{X}^t_{t-s}),\ \psi_1(\tilde{X}^t_{t-s}),\ \ldots,\ \psi_{q-1}(\tilde{X}^t_{t-s}))'$. Let ϕ be a given known measurable $r \times 1$ vector-valued function. Assume*

(a) $\hat{p}_1 \xrightarrow{p} p_0^*$, $\hat{p}_1 \xrightarrow{p} p_1^*$, $0 < p_0^*, p_1^* < 1$;

(b) (i) $\hat{M}_0 - p_0^{*^{-1}} M_1^* \xrightarrow{p} 0$, *where* $M_0^* \equiv T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_0]X_{\psi t}X_{\psi t}')$ *is* $O(1)$ *and uniformly non-singular*;

(ii) $\hat{M}_1 - p_1^{*^{-1}} M_1^* \xrightarrow{p} 0$, *where* $M_1^* \equiv T^{-1}\sum_{t=1}^{T} E(1(t \in \mathcal{T}_1]X_{\psi t}X_{\psi t}')$ *is* $O(1)$ *and uniformly non-singular;*

(c) (i) $\hat{K}_{\phi,0} - p_0^{*^{-1}} K_{\phi,0}^* \xrightarrow{p} 0$ *where* $K_{\phi,0}^* \equiv T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_0]X_{\psi t}\phi(V_t)')$ *is* $O(1)$;

(ii) $\hat{K}_{\phi,1} - p_1^{*^{-1}} K_{\phi,1}^* \xrightarrow{p} 0$ where $K_{\phi,1}^* \equiv T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_1]X_{\psi t}\phi(V_t)')$ *is* $O(1)$;

(d) *There exists a sequence of* $2qr \times 2qr$ *matrices* $\{V_T\}$, $O(1)$ *and uniformly non-singular, such that with* $\mathcal{T} \equiv \mathcal{T}_0 \cup \mathcal{T}_1$,

$$V_T^{-1/2}T^{-1/2}\sum_{t=1}^{T}\begin{bmatrix} 1[t \in \mathcal{T}](I \otimes X_{\psi t}) \\ 1[t \in \mathcal{T}_1](I \otimes X_{\psi t}) \end{bmatrix}\eta_t \xrightarrow{d} N(0, I),$$

where $\eta_t' \equiv \phi(\tilde{V}_t)' - X_{\psi t}'\alpha_{\phi,0}^* - \tilde{D}_t X_{\psi t}'(\alpha_{\phi,0}^* - \alpha_{\phi,1}^*)$, $\alpha_{\phi,0}^* \equiv M_0^{-1}K_{\phi,0}$, $\alpha_{\phi,1}^* \equiv M_1^{-1}K_{\phi,1}$, $\tilde{D}_t \equiv 1[t \in \mathcal{T}_1]$.

The conditions are stated at a relatively high level for conciseness, but can be verified for a useful spectrum of time-series processes. The conditions can also be considerably weakened, but with a corresponding increase in abstractness and complexity. The CIPP test statistic is now

$$\mathcal{W}_T \equiv T[\text{ vec } (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0})]'\hat{C}^{-1}[\text{ vec } (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0})],$$

where

$$\begin{aligned} \hat{C}_T &\equiv R\hat{M}^{-1}\hat{V}\hat{M}^{-1}R', \\ R &= [\mathbf{0}_{qr}, I_{qr}], \\ \hat{M} &= \begin{bmatrix} \hat{p}_0\hat{M}_0 + \hat{p}_1\hat{M}_1 & \hat{p}_1\hat{M}_1 \\ \hat{p}_1\hat{M}_1 & \hat{p}_1\hat{M}_1 \end{bmatrix}, \end{aligned}$$

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

$$\hat{V} = \tilde{T}^{-1} \sum_{t \in \mathcal{T}} \begin{bmatrix} I \otimes X_{\psi t} \\ \tilde{D}_t (I \otimes X_{\psi t}) \end{bmatrix} \hat{\eta}_t \hat{\eta}_t' \begin{bmatrix} I \otimes X_{\psi t} \\ \tilde{D}_t (I \otimes X_{\psi t}) \end{bmatrix}',$$

where $\tilde{T} \equiv T_0 + T_1$ and $\hat{\eta}_t' \equiv \phi(\tilde{V}_t)' - X_{\psi t}' \hat{\alpha}_{\phi,0} - \tilde{D}_t X_{\psi t}' (\hat{\alpha}_{\phi,1} - \hat{\alpha}_{\phi,0})$.

The following result establishes the properties of our test for dynamic CIPP.

Proposition 5.1:

(a) (i) *Given D.1-D.2(a-c),* $\hat{\alpha}_{\phi,0} - \alpha^*_{\phi,0} \xrightarrow{p} 0$, *and* $\hat{\alpha}_{\phi,1} - \alpha^*_{\phi,1} \xrightarrow{p} 0$, *where* $\alpha^*_{\phi,0} \equiv M_0^{*^{-1}} K^*_{\phi,0}$ *and* $\alpha^*_{\phi,1} \equiv M_1^{*^{-1}} K^*_{\phi,1}$ *are* $O(1)$.

(ii) *If in addition there exists* $a^*_{\phi,0}$ *such that* $E(\phi(V_t)'|1[t \in \mathcal{T}_0], \tilde{X}^t_{t-s}) = X'_{\psi t} a^*_{\phi,0}$ *and if*

$$\ddot{Z}^{\tau,t} \perp D^{\tau,t} \mid X^{\tau,t}, \qquad t = 1, 2, \ldots; \tau = 0, 1, \ldots,$$

then $\alpha^*_{\phi,0} = \alpha^*_{\phi,1} = a^*_{\phi,0}$.

(iii) *If in addition D.2(d) holds and* $\hat{V} - V_T \xrightarrow{p} 0$, *then* $\mathcal{W}_T \xrightarrow{d} \mathcal{X}^2_{qr}$.

(b) *Suppose D.*1-D.2(*a-c*) *hold such that* $\| \alpha^*_{\phi,0} - \alpha^*_{\phi,1} \| \geq \varepsilon > 0$ *and* $\hat{V} - V^*_T \xrightarrow{p} 0$, *with* $V^*_T = O(1)$ *and uniformly non-singular. Then for any sequence* $\{k_T\} = o(T)$, $P[\mathcal{W}_T > k_T] \to 1$.

Thus we reject (dynamic) CIPP at asymptotic level α whenever

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

where $\mathcal{X}^2_{qr,1-\alpha}$ is the $1-\alpha$ percentile of the $\mathcal{X}^2_{qr}$ distribution. When CIPP fails such that $\alpha^*_{\phi,0} \neq \alpha^*_{\phi,1}$, then a test with fixed α can detect this with probability approaching one. Further, the asymptotic level can be driven to zero, and, provided this is not too fast, power still approaches one. These results are entirely standard. Local power properties are also entirely standard. (Specialize results of White, 1994, ch. 8).

Now consider the consequences of rejecting or of failing to reject H_0. If one rejects H_0, then CIPP fails and the effects of interest cannot be consistently estimated using covariates X_t. This demonstrates the existence of causes (proxied by V_t) left out of account whose distribution differs between regimes and that may account for observed differences in outcomes between regimes.

Faced with rejection of H_0, one may append elements of V_t to W_t, augmenting the predictive proxies. (Note that our assumptions ensure validity of $\mathcal{V}$.) One may repeat the CIPP test using the augmented proxies, continuing to augment W_t until H_0 is no longer rejected or all available proxies have been utilized. Primary attention should be paid to finding proxies for the most important unobserved causes and/or for those causes whose distributions differ the most between regimes. Thus, in the cartel example finding a proxy for the price of an important raw material should be given high priority. This priority is enhanced if the price distribution appears to change dramatically between regimes, but it is diminished if its distribution is relatively stable.

Finally, consider the consequences of failing to reject CIPP. Then one has empirical evidence that accords with the consistent estimation of the effects of interest using covariates X_t. By constructing the test to achieve non-trivial power (by compelling choice of proxies and relevant transformations ϕ), such empirical evidence strengthens one's claim to have a useful measure of the effect of interest.

We note that even when CIPP is not rejected it may be helpful to augment W_t with the elements of V_t. The reason is that by including additional valid predictive proxies in the prediction equation, one reduces the variation of the prediction error, leading to more precise estimates of certain effects of interest.

For simplicity, we assume correct specification in Proposition 5.1(*ii*, *iii*); i.e., we assume $x'_\psi a^*_{\phi,0}$ exactly gives the conditional expectation of V_t given $\tilde{X}^t_{t-s} = x$ in regime 0. This can be straightforwardly weakened by letting $q \to \infty$ and implementing a fully nonparametric analysis.

6 Conclusion

In this paper we build on the labor econometrics and classical treatment effects literatures to provide a framework supporting causal concepts and methods for estimating effects of natural experiments operating over time in an explicitly dynamic time-series context. By exploiting the properties of valid predictive proxies, we also provide new tests for unconfoundedness in a dynamic setting.

As we have repeatedly emphasized, the work reported here is only a beginning. An array of exciting directions for future research presents itself. In particular, it is of interest to provide specific conditions ensuring the desirable properties of the various estimators outlined in Section 4, with particular attention paid to weighting methods exploiting the pseudo-propensity score and to the variety of relevant time-series data generating processes, such as stationary or co-integrated processes. Attention also needs to be paid to effects of interest beyond those specifically considered here, such as average effects across both treated and untreated cases or retrospective expected effects that condition on the full available history of the covariates. Application of the methods discussed here for analyzing the effects of natural experiments in panel data has particular relevance in the study of mergers of firms with geographically distinct operations. Exploration of the implications of the present conceptual framework for estimating the effects of recurring interventions, such as those considered by Romer and Romer (1989) and Angrist and Kuersteiner (2004) is also of interest.

Another direction in which the present results appear to extend readily is to variable treatments (let D_t be real-valued instead of {0,1}-valued) and to multiple treatments (let D_t be real vector-valued), based on suitable conditional independence conditions (see White, 2005b, ch. 6). Developing a fully nonparametric version of the CIPP test of Section 5 is a promising direction for future research. Most important of all, however, is to apply the methods described here to economic time-series data, as these methods offer the prospect of insights into the effects of natural experiments operating in time comparable to the causal insights now enjoyed by cross-section practitioners.

7 Mathematical Appendix

Proof of Proposition 3.1: 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

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

It follows straightforwardly that

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

as β_{01}^* satisfies the orthogonality conditions

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

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

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

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 3.1, it suffices to show that

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

We begin by decomposing

$$\begin{aligned} \beta^* \quad &\equiv \quad [(1-p_1)M_0 + p_1(M_1 - m_1m_1')]^{-1}[(1-p_1)L_0 + p_1(L_1 - m_1\mu_1)] \\ &= \quad [M_0 + \pi(M_1 - m_1m_1')]^{-1}[L_0 + \pi(L_1 - m_1\mu_1)], \end{aligned}$$

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

$$\begin{aligned} M_0 \quad &= \quad \begin{bmatrix} 1 & m_0'S' \\ Sm_0 & SM_0S' \end{bmatrix} \\ M_1 \quad &= \quad \begin{bmatrix} 1 & m_1'S' \\ Sm_1 & SM_1S' \end{bmatrix}, \end{aligned}$$

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

$$M_0 + \pi(M_1 - m_1m_1') = \begin{bmatrix} 1 & m_0'S' \\ Sm_0 & S(M_0 + \pi(M_1 - m_1m_1'))S' \end{bmatrix},$$

as the first row and column of $M_1 - m_1m_1'$ 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 3.1.

To obtain the second expression, it suffices to show that

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

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

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

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

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 3.1.

Proof of Proposition 3.2: Lemma 4.1 of Dawid (1979) ensures that $\ddot{Z}_t \perp D_t \mid (\tilde{Z}_t, W_t)$ implies $(\ddot{Z}_t, \tilde{Z}_t, W_t) \perp (D_t, \tilde{Z}_t, W_t) \mid (\tilde{Z}_t, W)$. It follows from Dawid (1979, Lemma 4.2) that $f(\ddot{Z}_t, \tilde{Z}_t, W_t) \perp g(D_t, \tilde{Z}_t, W_t) \mid (\tilde{Z}_t, W_t)$ 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 $g(d, \tilde{z}, \tilde{w}) = d$. It follows that

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

Proof of Theorem 4.1: Identical to that of 3.2 with minor notational changes.

Proof of Proposition 5.1. (i) It follows directly from Proposition 2.30 of White (2001) that $\hat{\alpha}_{\phi,0} - \alpha_{\phi,0}^* \xrightarrow{p} 0$ and $\hat{\alpha}_{\phi,1} - \alpha_{\phi,1}^* \xrightarrow{p} 0$, where

$$\alpha_{\phi,0}^* \equiv M_0^{*^{-1}} K_{\phi,0}^* \qquad \alpha_{\phi,1}^* = M_1^{*^{-1}} K_{\phi,1}^*.$$

(ii) We have

$$\begin{aligned} K_{\phi,0}^* &= T^{-1} \sum_{t=1}^{T} E(1[t \in \mathcal{T}_0] X_{\psi t} \phi(V_t)') \\ &= T^{-1} \sum_{t=1}^{T} E(E(1[t \in \mathcal{T}_0] X_{\psi t} \phi(V_t)' | t \in \mathcal{T}_0, \tilde{X}_{t-s}^t)) \\ &= T^{-1} \sum_{t=1}^{T} E(1[t \in \mathcal{T}_0] X_{\psi t} E(\phi(V_t)' | t \in \mathcal{T}_0, \tilde{X}_{t-s}^t)) \end{aligned}$$

$$
\begin{aligned}
&= T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_0] X_{\psi t} X'_{\psi t} a^*_{\phi,0}) \\
&= M_0^* a^*_{\psi,0},
\end{aligned}
$$

where the second to last equality holds because $E(\phi(V_t)'|t \in \mathcal{T}_0,\ \tilde{X}^t_{t-s}) = X'_{\psi t} a^*_{\phi 0}$. Consequently, $\alpha^*_{\phi,0} = a^*_{\psi,0}$. We also have

$$
\begin{aligned}
K^*_{\phi,1} &= T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_1] X_{\psi t} \phi(V_t)') \\
&= T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_1] X_{\psi t} E(\phi(V_t)'|t \in \mathcal{T}_1, \tilde{X}^t_{t-s})) \\
&= T^{-1}\sum_{t=1}^{T} E(1[t \in \mathcal{T}_1] X_{\psi t} E(\phi(V_t)'|t \in \mathcal{T}_0, \tilde{X}^t_{t-s})),
\end{aligned}
$$

because $\ddot{Z}^{\tau,t} \perp D^{\tau,t}|X^{\tau,t}$, $t = 1, 2, \ldots;\ \tau = 0, 1, \ldots$, ensures that

$$
\begin{aligned}
E(\phi(V_t)'|t &\in \mathcal{T}_1, \tilde{X}^t_{t-s} = \tilde{x}) = \int \phi(v(\ddot{z}, u, \tilde{x}))' dJ(u|\ddot{z}, \tilde{x})\, dG^t_{1,t-s}(\ddot{z}|\tilde{x}) \\
&= \int \phi(v(\ddot{z}, u, \tilde{x}))' dJ(u|\ddot{z}, \tilde{x})\, dG^t_{0,t-s}(\ddot{z}|\tilde{x}) \\
&= E(\phi(V_t)'|t \in \mathcal{T}_0, \tilde{X}^t_{t-s} = \tilde{x}) \\
&= x'_\psi a^*_{\psi,0}
\end{aligned}
$$

for all $t > s$. It follows that

$$
\begin{aligned}
K^*_{\phi,1} &= T^{-1}\sum_{t=1}^{T} E(1[t \in T_1] X_{\psi t} X'_{\psi t} a^*_{\phi,0}) \\
&= M_1^* a^*_{\phi,0},
\end{aligned}
$$

so that $\alpha^*_{\phi,1} = a^*_{\phi,0}$. (iii) The result follows immediately by application of Theorem 4.31 of White (2001). (iv) The conditions given ensure the consistency of $\hat{\alpha}_{\phi,0}$ for $\alpha^*_{\phi,0}$ and of $\hat{\alpha}_{\phi,1}$ for $\alpha^*_{\phi,1}$. The result follows by application of Theorem 8.16 of White (1994).

References

Angrist, J., 1998, Using social security data on military applicants to estimate the effect of voluntary military service on earnings. Econometrica 66, 249-288.

Angrist, J. and A. Krueger, 1999, Empirical strategies in labor economics, in: O. Ashenfelter and D. Card, (Eds.), Handbook of labor economics, Vol. 3A. Elsevier, Amsterdam, pp. 1277 - 1368.

Angrist, J. and G. Kuersteiner, 2004, Semiparametric causality tests using the policy propensity score, MIT Department of Economics Working Paper.

Barnow, B., G. Cain, and A. Goldberger, A., 1981, Selection on observables. Evaluation Studies Review Annual 5, 43-59.

Bauwens, L., H.P. Boswijk, and J.-P. Urbain, 2005. Causality and exogeneity in econometrics. Journal of Econometrics, in press.

Chen, X., 2005, Large sample sieve estimation of semi-nonparametric models. New York University C.V. Starr Center Working Paper.

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.

Corradi, V., W. Distaso, and N. Swanson, 2005, Predicting conditional confidence intervals of stochastic volatility via realized measures. Rutgers University Department of Economics Discussion Paper.

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.

Dawid, A.P., 2002, Influence diagrams for causal modeling and inference. International Statistical Review 70, 161-189.

Fisher, F., 1980, Multiple regression in legal proceedings. Columbia Law Review 85, 702-736.

Florens, J.-P. and D. Fougere, 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.

Gallant, A.R., and H. White, 1988. A unified theory of estimation and inference for nonlinear dynamic models. Blackwell, Oxford.

Goldberger, A., 1972, Structural equation methods in the social sciences. Econometrica 40, 979-1001.

Goldberger, A., 1991. A course in econometrics. Harvard University Press, Cambridge.

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. MacMillian, New York.

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, 1997, Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. Review of Economic Studies 64 605-654.

Heckman, J., H. Ichimura, and P. Todd, 1998, Matching as an econometric evaluation estimator. Review of Economic Studies 65, 261-294.

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. and G. Imbens, 2001, Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services & Outcomes Research 2, 259-278.

Hirano, K., G. Imbens, and G. Ridder, 2003, Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161-1189.

Hoover, K., 2001. Causality in macroeconomics. Cambridge University Press, Cambridge.

Horvitz, D. and J. Thompson, 1952, A generalization of sampling without replacement from a finite population. Journal of the American Statistical Association 47, 663-685.

Ing, C.-K., 2002, Multi-step prediction in autoregressive processes. Econometric Theory 19, 254-275.

Ing, C.-K., 2004, Selecting optimal multi-step predictors for autoregressive processes of unknown order. Annals of Statistics 32, 693-722.

Korenok, O. and N. Swanson, 2005, How sticky is sticky enough? A distributional and impulse response analysis of new Keynesian DSGE models. Rutgers University Department of Economics Discussion Paper.

Li, Q., J. Racine, and J. Wooldridge, 2004, Efficient estimation of average treatment effects in mixed categorical and continuous data. McMaster University Department of Economics Working Paper.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference.

Morgan Kaufman, San Mateo, CA.

Pearl, J.,1993a, Aspects of graphical methods connected with causality, in: Proceedings of the 49th session of the International Statistical Institute, pp. 391-401.

Pearl, J., 1993b, Comment: Graphical models, causality, and intervention. Statistical Science 8, 266-269.

Pearl, J., 1995, Causal diagrams for empirical research (with Discussion). Biometrika 82, 669-710.

Pearl, J., 2000. Causality. Cambridge University Press, Cambridge.

Reichenbach, H., 1956. The direction of time. University of California Press, Berkeley.

Robins, J., 1989, The control of confounding by intermediate variables. Statistics in Medicine 8, 679-701.

Robins, J., S. Greenland, and F.-C. Hu, 1999, Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. Journal of the American Statistical Association 94, 687-712.

Robins, J., M. Hernan, and B. Brumback, 2000, Marginal structural models and causal inference in epidemiology. Epidemiology 11, 550-560.

Romer, C. and D. Romer, 1989, Does monetary policy matter? NBER Macroeconomics Annual, 121-170.

Rosenbaum, P., 1984, The consequences of adjustment for a concomitant variable that has been affected by treatment. Journal of the Royal Statistical Society, Series A 147, 656-666.

Rosenbaum, P. and D. Rubin, 1983, The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41-55.

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.

Sims, C., 1972, Money, income, and causality. American Economic Review 62, 540-552.

Stinchcombe, M. and H. White, 1998, Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295-324.

Swanson, N. and R. Urbach, 2005, Simulation and prediction evidence on the usefulness of seasonal unit room models. Rutgers University Department of Economics discussion paper.

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.

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., 1994. Estimation, inference, and specification analysis. Cambridge University Press, New York.

White, H., 2001. Asymptotic theory for econometricians. Academic Press, New York.

White, H., 2005a, Estimating the effects of natural experiments, UCSD Department of Economics Working Paper.

White, H., 2005b. Causal, predictive, and explanatory modeling in economics. Oxford University Press, Oxford, forthcoming.