Granger Causality and Dynamic Structural Systems

Halbert White and Xun Lu*
Department of Economics
University of California, San Diego

June 26, 2008

Abstract

We analyze the relations between Granger (G) non-causality and a notion of structural causality arising naturally from a general nonseparable recursive dynamic structural system. Building on classical notions of G non-causality, we introduce interesting and natural extensions, namely weak G non-causality and retrospective weak G non-causality. We show that structural non-causality and certain (retrospective) conditional exogeneity conditions imply (retrospective) (weak) G non-causality. We strengthen structural causality to notions of (retrospective) strong causality and show that (retrospective) strong causality implies (retrospective) weak $G-$causality. We provide practical conditions and straightforward new methods for testing (retrospective) weak G non-causality, (retrospective) conditional exogeneity, and structural non-causality. Finally, we apply our methods to explore structural causality in industrial pricing, macroeconomics, and finance.

Keywords: Causality, Causality Testing, Granger Causality, Exogeneity, Conditional Exogeneity, Structural Systems.

JEL Classification Numbers: C30, C32, C51, E65

1 Introduction

In a celebrated paper, Granger (1969) introduced a notion now known as *Granger non-causality*, or, for brevity, "G non-causality." Since its introduction, it has been the focus of intense attention and interest, both theoretically and in applications. Specifically, G non-causality has often been used, either explicitly or implicitly, to gain insight into possible structural relations holding between the variables investigated. An example is

*Corresponding author email address: hwhite@ucsd.edu. We thank Clive Granger, Neil Ericcson, Karim Chalak, and seminar participants at UCSD's Economics Department for their helpful comments and suggestions.

Sims's (1972) seminal investigation of the causal relations between money and income. As Granger (1969) emphasizes, however, G non-causality is based purely on properties involving the predictability of particular time series of interest, and does not necessarily provide insight into whatever "true" causal relations may underlie the observed time series. Our goal here is thus to provide a direct link, previously missing, between G non-causality and a form of structural non-causality that emerges naturally from an explicit system of dynamic structural equations compatible with a wide range of economic data generating processes.

Not only does our analysis provide insight into situations where G non-causality is informative about structural non-causality and situations where it is not, it also provides explicit guidance as to how to properly apply G non-causality to obtain structural insight. Specifically, when testing G non-causality, certain variables in addition to the dependent variables (Y, say) and "potential G-causes" (D, say) play a crucial role in defining and testing G non-causality. For convenience, call these additional variables "covariates," and denote them S. Our results provide direct and specific guidance as to how the covariates should be chosen to ensure the desired link between G non-causality and structural non-causality. Specifically, this link holds when, among other things, the covariates S are chosen to be observable variables that structurally cause D or Y or are observable proxies for unobserved structural causes of either D or Y.

A further consequence of our analysis is the emergence of a variety of new and interesting natural extensions of the classical notions of G non-causality. Specifically, we introduce a notion of *weak* G non-causality that also is informative about structural non-causality but that makes use of a weaker information set. This weaker information set does not involve the entire past history of Y and thus leads to simpler tests. We also introduce notions of *retrospective* G non-causality; these extend both the classical and weak classical notions of G non-causality. Of particular interest is that retrospective (weak) G non-causality involves not just lags but also *leads* of the covariates. As we explain, these leads play a purely predictive role (in the back-casting sense); their presence thus does not violate the causal direction of time.

We obtain our results by making use of a system of general dynamic structural equations analyzed by White and Kennedy (2008). These systems permit data generating processes (DGPs) that may be nonlinear as well as nonseparable between observables and unobservables and that may generate stationary or nonstationary (e.g., cointegrated) processes. These systems support straightforward notions of structural causality.

Identification of structural effects is closely tied to certain conditional exogeneity assumptions, as discussed, for example, in White (2006a) and White and Kennedy (2008). Our formal results establish the relations between G non-causality, conditional exogeneity, and structural non-causality. Moreover, we show how new tests for G non-causality and conditional exogeneity can be combined to obtain new tests for structural non-causality.

The plan of the paper is as follows. In Section 2, we review classical notions of G non-causality. In Section 3 we introduce our dynamic DGP and notions of structural non-causality, and we introduce new notions of weak, retrospective, and retrospective weak G non-causality. We then provide our main results linking G non-causality, structural non-causality, and conditional exogeneity. In Section 4, we provide additional structure that leads to new and convenient methods for conducting tests of the various flavors of G non-causality. Section 5 discusses new tests for (retrospective) conditional exogeneity and a pure test for structural non-causality, based on tests for (retrospective) (weak) G non-causality and (retrospective) conditional exogeneity. Section 6 contains illustrations of our methods, involving applications to gasoline and oil prices (as in White and Kennedy, 2008); monetary policy and industrial production (as in Romer and Romer, 1989 and Angrist and Kuersteiner, 2004); and stock returns and macroeconomic announcements (as in Chen, Roll, and Ross, 1986 and Flannery and Protopapadakis, 2002). Section 7 contains a summary and concluding remarks.

2 Granger Non-Causality

Granger (1969) defined G non-causality in terms of conditional expectations. Granger and Newbold (1986) extend the concept to a definition in terms of conditional distributions. Here, we work with the latter approach. In our discussion to follow, we adapt the notation of these sources, but otherwise preserve the conceptual content.

For any sequence of random vectors $\{Y_t\}$, let $Y^t \equiv (Y_0, ..., Y_t)$ denote the "$t-$history" of the sequence, and let $\sigma(Y^t)$ denote the sigma-field generated by Y^t. That is, $\sigma(Y^t)$ is the "information set" associated with Y^t. Let $\{D_t, S_t, Y_t\}$ be a sequence of random vectors. Granger and Newbold (1986) say that D_t does not G-cause Y_t with respect to $\sigma(D^t, S^t, Y^t)$ if for all $t \geq 0$,

$$F_{t+k}(\ \cdot \mid D^t, S^t, Y^t) = F_{t+k}(\ \cdot \mid S^t, Y^t), \quad k = 1, 2, ..., \tag{1}$$

where $F_{t+k}(\ \cdot \mid D^t, S^t, Y^t)$ denotes the conditional distribution function of Y_{t+k} given D^t, S^t, Y^t, and $F_{t+k}(\cdot \mid S^t, Y^t)$ denotes the conditional distribution function of Y_{t+k} given S^t, Y^t.

In the special case in which $\sigma(D^t, S^t, Y^t) = \sigma(\mathcal{U}^t)$, where $\mathcal{U}_t$ is the "universal" random vector, so that $\sigma(\mathcal{U}^t)$ contains all information in the universe up to time t, Granger and Newbold drop the reference to the information set $\sigma(D^t, S^t, Y^t)$. When $\sigma(D^t, S^t, Y^t) \subset \sigma(\mathcal{U}^t)$, Granger and Newbold also say that D_t is not a "prima facie" $G-$cause for Y_t.

If eq.(1) does not hold, Granger and Newbold say that D_t does $G-$cause Y_t with respect to $\sigma(D^t, S^t, Y^t)$ or that D_t is a prima facie $G-$cause for Y_t. If $\sigma(D^t, S^t, Y^t) = \sigma(\mathcal{U}^t)$, the "prima facie" qualifier is dropped. As Granger (1969) and Granger and Newbold (1986) further note, however, the use of the universal information set is not practical, so the typical case is that in which $\sigma(D^t, S^t, Y^t) \subset \sigma(\mathcal{U}^t)$.

Granger and Newbold (1986, p.221) caution that

> Not everyone would agree that causation is the correct term to use for this situation, but we shall continue to do so as it is both simple and clearly defined.

As Granger and Newbold (1986, p.222) further note,

> It has been suggested, for example, that causation can only be accepted if the empirical evidence is associated with a clear and convincing theory of how the cause produces the effect. If this viewpoint is accepted, then "smoking causes cancer" would not be accepted.

Granger and Newbold do not endorse the requirement of a clear and convincing theory of how causes produce effects. At the extreme, this suggests that notions of $G-$causality can involve any variables whatsoever, regardless of underlying structural relationships. Indeed, as Granger (1969, p.430) notes, "The definition of causality used above is based entirely on the predictability of some series." Knowledge about underlying structural relationships may thus be helpful in investigating $G-$causality, but is by no means necessary.

Of particular interest here is the reciprocal fact, established in what follows, that knowledge about $G-$causality may be helpful in investigating structural relationships. This idea is often implicit in empirical tests of $G-$causality; our results below make the relations between $G-$causality and structural causality explicit and precise.

As noted by Florens and Mouchart (1982) and Florens and Fougère (1995), G non-causality is a conditional independence requirement. Following Dawid (1979) (D), we write $\mathcal{X} \perp \mathcal{Y} \mid \mathcal{Z}$ when $\mathcal{X}$ and $\mathcal{Y}$ are independent given $\mathcal{Z}$. Letting $\mathbb{N}^+ \equiv \{1, 2, ...\}$ and $\mathbb{N} \equiv \{0\} \cup \mathbb{N}^+$, we formally define Granger non-causality as follows:

Definition 2.1 *Let* $\{D_t, S_t, Y_t\}$ *be a sequence of random vectors, and let* $\mathbb{K} \subset \mathbb{N}^+$. *Suppose that*

$$Y_{t+k} \perp D^t \mid Y^t, S^t \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{K}. \tag{2}$$

If $\mathbb{K} = \{1\}$, *then we say that* D does not $G-$cause Y with respect to S. *Otherwise, we say that* D $G-$causes Y with respect to S. *If* $\mathbb{K} = \mathbb{N}^+$, *then we say that* D does not G^+-cause Y with respect to S. *Otherwise, we say that* D G^+-causes Y with respect to S.

The key idea is that D^t provides no information useful in predicting Y_{t+k} beyond the information contained in the histories S^t and Y^t for the given values of k. The specification $k = 1$ is implicit in Granger (1969) and explicit in Granger (1980, 1988). We thus apply the standard terminology ("$G-$cause") for this case. The case with $k \geq 1$ appears in Granger and Newbold (1986). The superscript in G^+ is intended to suggest that the condition holds for all $k \in \mathbb{N}^+$. Other choices for $\mathbb{K}$ are possible, but these will not play a role here.

Note that G^+ non-causality implies G non-causality. Thus if D $G-$causes Y with respect to S, we also have that D G^+-causes Y with respect to S.

3 Granger Causality and Structural Causality

3.1 A dynamic DGP and structural causality

We now specify a data generating process (DGP) as a particular dynamic structural system of equations. This is a version of the structure analyzed in White and Kennedy (2008). As White and Chalak (2007a) (WC) discuss, such systems support clear structural definitions of causal effects.

Assumption A.1 (a) Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, on which are defined the random vectors D_0, V_0, W_0, Y_0, and the stochastic process $\{Z_t\}$, where D_0, V_0, W_0, Y_0, and Z_t take values in $\mathbb{R}^{k_d}, \mathbb{R}^{k_v}, \mathbb{R}^{k_w}, \mathbb{R}^{k_y}$, and $\mathbb{R}^{k_z}$ respectively, where k_v and k_z are countably valued integers and k_d, k_w, and k_y are finite integers, $k_d, k_y > 0$. Further, let

$\{D_t, V_t, W_t, Y_t\}$ be a sequence of random vectors generated as

$$
\begin{aligned}
V_{t+1} &\overset{c}{=} b_{0,t+1}(V^t, Z^t) \\
W_{t+1} &\overset{c}{=} b_{1,t+1}(W^t, V^t, Z^t) \\
D_{t+1} &\overset{c}{=} b_{2,t+1}(D^t, W^t, V^t, Z^t) \\
Y_{t+1} &\overset{c}{=} q_{t+1}(Y^t, D^t, V^t, Z^t), \qquad t = 0, 1, ...,
\end{aligned}
\tag{3}
$$

where $b_{0,t+1}, b_{1,t+1}, b_{2,t+1}$, and q_{t+1} are unknown measurable functions of dimension k_v, k_w, k_d, and k_y, respectively.

(b) For $t = 0, 1, ...,$ $V_t \equiv (\tilde{V}_t, \ddot{V}_t)$ and $Z_t \equiv (\tilde{Z}_t, \ddot{Z}_t)$, where $\tilde{V}_t$ and $\tilde{Z}_t$ take values in $\mathbb{R}^{k_{\tilde{v}}}$ and $\mathbb{R}^{k_{\tilde{z}}}$ respectively, and $k_{\tilde{v}}$ and $k_{\tilde{z}}$ are finite integers. Realizations of $Y_t, D_t, \tilde{V}_t, W_t$, and $\tilde{Z}_t$ are observed; realizations of $\ddot{V}_t$ and $\ddot{Z}_t$ are not observed.

We view this system as representing the causal structure holding among the various components of the system. We follow WC and Chalak and White (2007c) in using the notation $\overset{c}{=}$ to emphasize that the structural equations (3) represent asymmetric causal links (Goldberger, 1972, p.979), in which manipulations of elements of y^t, d^t, v^t, z^t result in potentially differing values for y_{t+1}, as in Strotz and Wold (1960) and Fisher (1966, 1970). Leading examples of such structures are those that arise from the dynamic optimization behavior of economic agents and/or interactions among such agents. Chow (1997) provides numerous examples.

Observe that this dynamic structure is general, in that the structural relations may be nonlinear and non-monotonic in their arguments and non-separable between observables and unobservables. The unobservables may be countably infinite in number. Finally, this system may generate stationary processes, non-stationary processes, or both.

This dynamic structure is a mild restriction of that given by White and Kennedy (2008). There, "instantaneous causation" is permitted (but not required) by letting, e.g., D_{t+1}, V_{t+1}, and Z_{t+1} appear as arguments of q_{t+1}. Here, we suppress this, in keeping with the spirit of Granger's (1969, 1988) and Granger and Newbold's (1986) views on instantaneous causation. For example, Granger and Newbold (1986, p.221) note that

> Whether all IC (instantaneous causality) can be explained in terms of data inadequacies is unclear... . However, it is clear that it is not possible, in general, to differentiate between instantaneous causation in either direction

and instantaneous feedback. Thus, the idea of instantaneous causality is of little or no practical value.

The vector Y_{t+1} represents responses of interest, and we focus our attention on the effects of D^t on Y_{t+1}. Because of the dynamics (lagged Y_t's) appearing in q_t, these effects can propagate through time. To accommodate this, we work with an equivalent implicit dynamic representation of the DGP. By recursive substitution, we have

$$\begin{aligned} V_{t+1} &\overset{c}{=} c_{0,t+1}(V_0, Z^t) \\ W_{t+1} &\overset{c}{=} c_{1,t+1}(W_0, V^t, Z^t) \\ D_{t+1} &\overset{c}{=} c_{2,t+1}(D_0, W^t, V^t, Z^t) \\ Y_{t+1} &\overset{c}{=} r_{t+1}(Y_0, D^t, V^t, Z^t), \qquad t = 0, 1, ..., \end{aligned} \tag{4}$$

where $c_{0,t}, c_{1,t}, c_{2,t}$, and r_t are unknown functions. Thus, $\{Y_t, D_t, W_t, V_t, Z_t\}$ is determined entirely by $\tilde{\mathcal{S}} \equiv \{(\Omega, \mathcal{F}, P), (D_0, V_0, W_0, Y_0, \{Z_t\}; \{b_t, q_t\})\}$, where $\{b_t\} \equiv \{b_{0,t}, b_{1,t}, b_{2,t}\}$, or equivalently by $\mathcal{S} \equiv \{(\Omega, \mathcal{F}, P), (D_0, V_0, W_0, Y_0, \{Z_t\}; \{c_t, r_t\})\}$, where $\{c_t\} \equiv \{c_{0,t}, c_{1,t}, c_{2,t}\}$. In the nomenclature of WC, $\mathcal{S}$ is a *canonical recursive settable system.* For such systems, WC introduce the following structurally based definition of causality.

Definition 3.1 *Suppose that for given t and all admissible values of y_0, v^t, and z^t, the function $d^t \to r_{t+1}(y_0, d^t, v^t, z^t)$ is constant in d^t. Then we say that D^t* does not cause Y_{t+1} with respect to $\mathcal{S}$*, and we write $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$. Otherwise, we say that D^t* causes Y_{t+1} with respect to $\mathcal{S}$*, and we write $D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$.*

We state the definition in terms of D^t, as this is the case we pay most attention to. A similar definition holds for the individual elements of D^t. Because of the structural basis for this definition of causality, we refer to it as "structural causality" to distinguish it from other notions of causality. This notion is broadly consistent with the causal notions implicit in the work of the Cowles Commission, and in particular with those explicitly articulated by Heckman (e.g., Heckman, 2008).

Given our focus on the effects of D^t, we call D_t "causes of interest." The histories V^t and Z^t contain causes of Y_{t+1} whose effects are not of primary interest; we call V_t and Z_t "ancillary causes." The variables Z_t represent "fundamental" variables, that is, variables that are structurally exogenous, in that they are not determined within the system.

According to A.1(b), we observe Y_t and D_t. We also observe the finite subvectors $\tilde{V}_t$ and $\tilde{Z}_t$ of V_t and Z_t; we call these "observed ancillary causes." The observable variables W_t act as proxies for the countably dimensioned vector of "unobserved ancillary causes," $U_t \equiv (\ddot{V}_t, \ddot{Z}_t)$. Observe that W^t may (but need not) directly determine D_{t+1}, but W^t does not directly determine Y_{t+1}.

3.2 Structural causality and classical G-causality

Our first formal result forges a link between structural non-causality and the classical notions of G or G^+ non-causality.

Proposition 3.2 *Suppose A.1(a) holds.*

(a) Then

$$Y_{t+k} \perp D^t \mid D_0, W^{t-1}, V^{t-1}, Z^{t-1} \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{N}^+; \quad \text{and} \tag{5}$$

$$Y_{t+k} \perp D^t \mid Y^t, D_0, W^{t-1}, V^{t-1}, Z^{t-1} \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{N}^+. \tag{6}$$

(b) Suppose in addition that $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$ for all $t \in \mathbb{N}$. Then

$$Y_{t+k} \perp D^t \mid Y_0, V^t, Z^{t+k-1} \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{N}^+; \quad \text{and} \tag{7}$$

$$Y_{t+k} \perp D^t \mid Y^t, V^t, Z^{t+k-1} \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{N}^+. \tag{8}$$

Part (a) shows that the specific choice $S^t = (D_0, W^{t-1}, V^{t-1}, Z^{t-1})$ guarantees G and G^+ non-causality, *even when D^t structurally causes Y_{t+1}*. Eqs.(5) and (6) follow directly from D, lemmas 4.1 and 4.2, as A.1(a) ensures that D^t is measurable with respect to both $\sigma(D_0, W^{t-1}, V^{t-1}, Z^{t-1})$ and $\sigma(D_0, Y^t, W^{t-1}, V^{t-1}, Z^{t-1})$. Thus, the "right" information set can break any link that might have been thought to exist from structural non-causality to Granger non-causality.

In part (b), structural non-causality does imply G non-causality, but now with $S^t = (V^t, Z^t)$. This result holds because if $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$, then Y_{t+k} is measurable with respect to both $\sigma(Y_0, V^t, Z^{t+k-1})$ and $\sigma(Y^t, V^t, Z^{t+k-1})$, straightforwardly delivering eqs.(7) and (8).

Although eq.(8) resembles G^+ non-causality, it is distinct, as here Z^{t+k-1} for $k > 1$ appears in the conditioning set. Thus, we generally can *not* claim that structural non-causality implies G^+ non-causality for this choice of S^t, because under A.1(a) we can easily have $Z_{t+1}^{t+k-1} \not\perp D^t \mid Y^t, V^t, Z^t$, and it can then easily happen that $Y_{t+k} = r_{t+k}\left(Y_0, V^{t+k-1}, Z^{t+k-1}\right) \not\perp D^t \mid Y^t, V^t, Z^t$.

This latter result, in which we have G non-causality, but not G^+ non-causality, is the first of several indications we encounter suggesting that G^+ non-causality may be overly restrictive, and that G non-causality is the more fruitful concept in the presence of explicit dynamic structure.

3.3 Weak G-causality and conditional exogeneity

Given A.1(b), Proposition 3.2 involves unobservables, so it is not of immediate practical value. For use in applications, we seek results that involve only observable random variables. To explore the possibilities afforded by the structure in A.1(a), we introduce some useful extensions of G and G^+ non-causality. First, we define *weak* G and G^+ non-causality:

Definition 3.3 *Let $\{D_t, S_t, Y_t\}$ be a sequence of random variables, and let $\mathbb{K} \subset \mathbb{N}^+$. Suppose that*

$$Y_{t+k} \perp D^t \mid Y_0, S^t \quad \text{for all } t \in \mathbb{N} \text{ and all } k \in \mathbb{K}. \tag{9}$$

If $\mathbb{K} = \{1\}$, then we say that D does not weakly *$G-$cause Y with respect to S. Otherwise, we say that D* weakly *$G-$causes Y with respect to S. If $\mathbb{K} = \mathbb{N}^+$, then we say that D does not* weakly *G^+-cause Y with respect to S. Otherwise, we say that D* weakly *G^+-causes Y with respect to S.*

Weak G (G^+) non-causality differs from classical G (G^+) non-causality in that instead of conditioning on Y^t, S^t, we condition on the weaker information set generated by Y_0, S^t. This concept appears not to have been previously introduced, perhaps because it could have significant disadvantages in a purely predictive context. Nevertheless, this restriction turns out to be a natural one, once the general dynamic structure of A.1 is specified. Moreover, because the entire past history Y^t is no longer involved, this permits more parsimonious tests and may thus offer strong practical advantages in empirical applications.

As suggested above, our nomenclature is driven by the fact that the information set generated by Y_0, S^t is "weaker" (less informative) than that generated by Y^t, S^t. Nevertheless, there is no necessary relation between weak and classical Granger non-causality concepts. Without further conditions, neither is necessary nor sufficient for the other.

White (2006a) shows how certain conditional exogeneity restrictions permit the identification of a variety of causal effects of interest in dynamic structural systems. Although

White (2006a) explicitly considers only binary-valued D_t, WC and White and Kennedy (2008) show that suitable conditional exogeneity assumptions identify effects of interest generally. As we show next, conditional exogeneity also implies links between structural non-causality and various modes of Granger non-causality involving only observables. To state a conditional exogeneity assumption, we define the *covariates* $X_t \equiv (\tilde{V}_t, W_t, \tilde{Z}_t)$. These are observable under A.1(b).

Assumption A.2 (a) $D^t \perp U^t \mid Y_0, X^t$, $t = 0, 1, 2, ...$.

When A.2(a) holds, we say that D^t is *conditionally exogenous with respect to* U^t *given* (Y_0, X^t), $t = 0, 1, 2, ...$. For brevity, we may just say that D^t is *conditionally exogenous*. See White (2006a) and White and Kennedy (2008) for discussion of this concept. We remark that A.1(a) ensures that A.2(a) is equivalent to $D^t \perp (\ddot{V}_0, \ddot{Z}^t) \mid Y_0, X^t$.

We now state a main result formally linking structural non-causality to different modes of Granger non-causality.

Proposition 3.4 *Suppose Assumption A.1(a) holds and that* $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$ *for all* $t \in \mathbb{N}$. *If Assumption A.2(a) also holds (conditional exogeneity), then* D *does not (weakly)* $G-$*cause* Y *with respect to* X.

This key result provides a direct link between structural non-causality and G non-causality with $S^t = X^t$. Proposition 3.4 implies that if one tests and rejects the null hypothesis of G non-causality, then we have either structural causality ($D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$) or the failure of conditional exogeneity or both. As indicated by our parenthetical reference to weak G non-causality, rejection of weak G non-causality also suffices to reject structural non-causality or conditional exogeneity.

If conditional exogeneity is taken as given (that is, if one is willing to assume that causal effects can be identified), then the rejection of (weak) G non-causality implies structural causality. If structural causality is taken as given, then the rejection of (weak) G non-causality implies rejection of conditional exogeneity. In the absence of conditional exogeneity, causal effects are generally not identified, as discussed in WC.

These conditions imply neither weak nor classical G^+ non-causality, for reasons analogous to those given following Proposition 3.2.

3.4 Retrospective weak G-causality and retrospective conditional exogeneity

White and Kennedy (2008) consider retrospective expected effects, in which expected counterfactual responses are conditioned on all information available at the present, time T. Identification of such retrospective effects is ensured by a *retrospective* conditional exogeneity condition, which we formally impose as

Assumption A.2 (b) For given $T \in \mathbb{N}^+$, $D^t \perp U^t \mid Y_0, X^T$, $t = 0, 1, ..., T$.

Whereas A.2(a) conditions only on past covariates relative to time $t+1$, A.2(b) conditions not only on past covariates but also present and future covariates, relative to $t+1$. As White and Kennedy (2008) discuss in detail, this does not violate the causal order of time, because the causal structure of A.1(a) respects the causal order of time. Indeed, the conditioning in A.2(a) and A.2(b) is purely a *stochastic* operation, with no causal content. The role of the conditioning variables, and in particular any "future" covariates among the conditioning variables, is solely predictive. "Future" covariates are predictive in the back-casting sense. Augmenting the covariates as in A.2(b) can be especially useful, as the additional information in X^T relative to that in X^t may enable A.2(b) to hold when A.2(a) fails.

Corresponding to retrospective conditional exogeneity are the following retrospective Granger non-causality conditions:

Definition 3.5 *Let $\{D_t, S_t, Y_t\}$ be a sequence of random variables. For a given $T \in \mathbb{N}^+$ and integer $\theta \leq T$, let $\mathbb{K}_\theta := \{1, ..., \theta\}$.*

(a) Suppose that

$$Y_{t+k} \perp D^t \mid Y^t, S^T \text{ for all } 0 \leq t \leq T - \theta \text{ and all } k \in \mathbb{K}_\theta.$$

Then we say that D does not retrospectively *$G_\theta-$cause Y with respect to S. Otherwise, we say that D* retrospectively *$G_\theta-$causes Y with respect to S.*

(b) Suppose that

$$Y_{t+k} \perp D^t \mid Y_0, S^T \text{ for all } 0 \leq t \leq T - \theta \text{ and all } k \in \mathbb{K}_\theta.$$

Then we say that D does not retrospectively *weakly $G_\theta-$cause Y with respect to S. Otherwise, we say that D* retrospectively *weakly $G_\theta-$causes Y with respect to S.*

The case where $\theta = 1$ is that most relevant here. For convenience, we thus refer to retrospective (weak) G_1 (non-) causality simply as retrospective (weak) G (non-) causality.

Just as for conditional exogeneity, there is no direct relation between weak and retrospective weak Granger (non-) causality. Neither is necessary nor sufficient for the other. Nor are there any necessary relations between retrospective and classical Granger (non-) causality.

We now state a main result formally linking structural non-causality to retrospective (weak) Granger non-causality, parallel to Proposition 3.4.

Proposition 3.6 *Suppose Assumption A.1(a) holds and that $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$ for all $t \in \mathbb{N}$. If Assumption A.2(b) also holds (retrospective conditional exogeneity), then for the given T, D does not retrospectively (weakly) $G-$cause Y with respect to X.*

The structure wrovided by A.1(a) and A.2(b) thus implies both retrospective agd retrospective weak G non-causality with $S^T = X^T$ when D^t does not structurally cause Y_{t+1}.

3.5 Some converse results

So far, our results establish that given conditional exogeneity, structural non-causality implies various forms of Granger non-causality. Our next result is a form of converse, establishing that weak G non-causality implies a form of structural non-causality. To establish this, we do not require conditional exogeneity; instead, we use a strengthened version of structural causality.

Assumption A.3 For given $t \in \mathbb{N}$, suppose

(a) there exist measurable sets B_Y, B_0, B_D, and B_X such that:

(i)

$$P[Y_{t+1} \in B_Y, Y_0 \in B_0, D^t \in B_D, X^t \in B_X] > 0;$$

(ii)

$$P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X] < 1; \quad \textit{and}$$

(iii) with $B_U(d^t, y_0, x^t) \equiv \text{supp}(U^t \mid D^t = d^t, Y_0 = y_0, X^t = x^t)$, for all $d^t \notin B_D$, $y_0 \in B_0$, and $x^t \in B_X$, and all $u^t \in B_U(d^t, y_0, x^t)$

$$r_{t+1}(y_0, d^t, \tilde{v}^t, \tilde{z}^t, \ddot{v}^t, \ddot{z}^t) \notin B_Y.$$

Observe that the requirement of A.3(a.i),

$$P[Y_{t+1} \in B_Y, Y_0 \in B_0, D^t \in B_D, X^t \in B_X] > 0,$$

implies

$$P[Y_0 \in B_0, D^t \in B_D, X^t \in B_X] > 0 \quad \text{and} \quad P[Y_0 \in B_0, X^t \in B_X] > 0.$$

This in turn implies that

$$\begin{aligned} P[D^t &\in B_D \mid Y_0 \in B_0, X^t \in B_X] \\ &= P[Y_0 \in B_0, D^t \in B_D, X^t \in B_X]/P[Y_0 \in B_0, X^t \in B_X] \\ &> 0. \end{aligned}$$

Assumptions A.3(a.i) and A.3(a.ii) ensure $0 < P[D^t \in B_D|Y_0 \in B_0, X^t \in B_X] < 1$.

Assumption A.3(a) is stronger than assuming that $D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$, because A.3(a) implies $D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$, but $D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$ does not ensure A.3(a). In part, the idea is that the values of the ancillary causes ensuring $D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$ may occur on a set with probability zero. If so, we will not be able to detect structural causality in data. Assumption A.3(a) rules this out. But Assumption A.3(a) goes further; it ensures that there are response values (those in the set B_Y) that can only be reached by variation in D^t, for some sets of conditioning values of Y_0 and X^t (B_0 and B_X).

In fact, A.3(a) is quite a strong assumption. Consider, for example, the simple case in which $Y_t = D_t + U_t$, where D_t has unbounded support. Then A.3(a) can hold if the support of U_t is conditionally bounded, but not otherwise. Because of its strength, this condition may be plausible in some applications but not others. We offer A.3(a) and the results that follow as a first step in investigating conditions ensuring that structural causality implies $G-$causality. Despite its strength, A.3(a) nevertheless affords useful insights, as discussed immediately below.

We provide a convenient way of referring to Assumption A.3(a) using the following definition.

Definition 3.7 *(a) Suppose A.1 and A.3(a) hold. Then we say that D^t* strongly *causes Y_{t+1} with respect to $\mathcal{S}$, and we write $D^t \overset{s}{\Rightarrow}_{\mathcal{S}} Y_{t+1}$. Otherwise, we say that D^t* does not strongly *cause Y_{t+1} with respect to $\mathcal{S}$, and we write $D^t \overset{s}{\not\Rightarrow}_{\mathcal{S}} Y_{t+1}$.*

We can now state a form of converse to Proposition 3.4.

Theorem 3.8 *(a) Let Assumptions A.1 and A.3 (a) hold, i.e.,* $D^t \overset{s}{\Rightarrow}_{\mathcal{S}} Y_{t+1}$. *Then* D *weakly G-causes* Y *with respect to* X.

Note that we do not necessarily have $Y_{t+k} \not\perp D^t \mid Y_0, X^t$ for $k > 1$. That is, strong structural causality does not imply weak G^+-causality.

Theorem 3.8 implies that if one tests and *fails* to reject the null hypothesis of weak G non-causality (e.g., using the methods discussed below), then one has evidence consistent with the absence of strong structural causality. The evidence is also consistent with structural causality ($D^t \Rightarrow_{\mathcal{S}} Y_{t+1}$), but if this is present, it is not strong enough to satisfy Assumption A.3(a). Note that this result neither makes assumptions nor draws conclusions about conditional exogeneity.

Analogous conclusions follow under a retrospective version of Assumption A.3(a):

Assumption A.3 (b) For given $t \in \mathbb{N}$, $T \in \mathbb{N}^+$ with $t \leq T$, suppose conditions A.3(a.i)-(a.iii) hold with X^T and x^T replacing X^t and x^t respectively.

Definition 3.7 *(b) Suppose A1 and A.3(b) hold. Then we say that* D^t retrospectively strongly causes Y_{t+1} with respect to $\mathcal{S}$*, and we write* $D^t \overset{r,s}{\Rightarrow}_{\mathcal{S}} Y_{t+1}$. *Otherwise, we say that* D^t does not retrospectively strongly cause Y_{t+1} with respect to $\mathcal{S}$*, and we write* $D^t \overset{r,s}{\not\Rightarrow}_{\mathcal{S}} Y_{t+1}$.

Theorem 3.8 *(b) Let Assumptions A.1 and A.3 (b) hold, i.e.,* $D^t \overset{r,s}{\Rightarrow}_{\mathcal{S}} Y_{t+1}$. *Then* D *retrospectively weakly G-causes* Y *with respect to* X.

4 Practical Conditions for Weak G Non-Causality

Our various notions of Granger non-causality all involve the histories D^t and X^t or X^T, and the classical notions further involve the history Y^t. Testing these notions is challenging in pure time-series data, where there is just one data history $\{Y_t, D_t, X_t\}_{t=0}^T$. Nevertheless, imposing some suitable structure makes it possible to construct practical tests in a straightforward manner. Specifically, White and Kennedy (2008) impose the following condition.

Assumption A.1 (a′) Assumption A.1(a) holds with $Y_{t+1} \overset{c}{=} q_{t+1}(Y_t, D_t, V_t, Z_t)$ for all $t \in \mathbb{N}$.

Assumption A.1(a$'$) imposes "first order" dynamic structure on the data generating process for $\{Y_t\}$. Analogous results hold with any finite order dynamic structure, but we consider first order structure for concreteness and clarity.

To state the next assumption, let $dF_{t+1,t}(\cdot \mid y_t, d_t, x^t)$ denote the conditional density of Y_{t+1} given $Y_t = y_t, D_t = d_t$, and $X^t = x^t$, and let $dF_{-\tau}(\cdot \mid y_t, d_t, x^t_{t-\tau})$ denote the conditional density of Y_{t+1} given $Y_t = y_t, D_t = d_t$, and $X^t_{t-\tau} = x^t_{t-\tau}$, where $X^t_{t-\tau} \equiv (X_{t-\tau}, ..., X_t)$ is the one-sided $(\tau-)$ near history of X_t. Similarly, let $dF_{t+1,T}(\cdot \mid y_t, d_t, x^T)$ denote the conditional density of Y_{t+1} given $Y_t = y_t, D_t = d_t$, and $X^T = x^T$, and let $dF_\tau(\cdot \mid y_t, d_t, x^{t+\tau}_{t-\tau})$ denote the conditional density of Y_{t+1} given $Y_t = y_t, D_t = d_t$, and $X^{t+\tau}_{t-\tau} = x^{t+\tau}_{t-\tau}$, where $X^{t+\tau}_{t-\tau} \equiv (X_{t-\tau}, ..., X_{t+\tau})$ is the two-sided $(\tau-)$ near history of X_t.

Assumption A.4

(a) There exist $\tau \in \mathbb{N}$ and a conditional density $dF_{-\tau}$ such that for all $t \in \mathbb{N}$ and for all argument values

$$dF_{t+1,t}(y_{t+1}|y_t, d_t, x^t) = dF_{-\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau}).$$

(b) There exist $\tau \in \mathbb{N}$ and a conditional density dF_τ such that for given $T \in \mathbb{N}^+$, all integers $t \leq T - \tau$, and for all argument values

$$dF_{t+1,T}(y_{t+1}|y_t, d_t, x^T) = dF_\tau(y_{t+1}|y_t, d_t, x^{t+\tau}_{t-\tau}).$$

Assumption A.4 combines a memory condition with a conditional stationarity assumption. The memory condition says that only the near history $X^t_{t-\tau}$ $(X^{t+\tau}_{t-\tau})$ of the covariates is useful in predicting the response Y_{t+1}, given Y_t and D_t. The fact that $dF_{-\tau}$ (dF_τ) does not have a subscript t expresses the assumption that the conditional density is the same for all t, thereby imposing conditional stationarity. Observe that conditional stationarity does not rule out integrated or cointegrated processes.

For the retrospective case, White and Kennedy (2008) establish a recursive representation of $dF_{t+1,T}(y_{t+1} \mid y_0, d^t, x^T)$. We state a version of this in our next result (Proposition 4.1(b)). To obtain a similar representation for $dF_{t+1,t}(y_{t+1} \mid y_0, d^t, x^t)$, we impose one further assumption.

Assumption A.5 $Y_t \perp X_t \mid Y_0, D^{t-1}, X^{t-1}$, $t = 0, 1, 2, ...$.

This assumption is not implied by our prior assumptions, but it is nevertheless plausible given that the elements of X_t do not enter the structural equation for Y_t under A.1(a$'$).

Proposition 4.1

(a) Suppose Assumptions A.1(a′), A.2(a), A.4(a), and A.5 hold. Then

$$\begin{aligned} dF_{1,0}(y_1|y_0, d^0, x^0) &= dF_{-\tau}(y_1|y_0, d_0, x^0_{-\tau}) \qquad (t=0) \\ dF_{t+1,t}(y_{t+1}|y_0, d^t, x^t) &= \int dF_{-\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau})\ dF_{t,t-1}(y_t|y_0, d^{t-1}, x^{t-1}) \\ & \qquad t = 1, 2, \ldots . \end{aligned}$$

(b) Suppose Assumptions A.1(a′), A.2(b), and A.4(b) hold. Then

$$\begin{aligned} dF_{1,T}(y_1|y_0, d^0, x^T) &= dF_{\tau}(y_1|y_0, d_0, x^{\tau}_{-\tau}) \qquad (t=0) \\ dF_{t+1,T}(y_{t+1}|y_0, d^t, x^T) &= \int dF_{\tau}(y_{t+1}|y_t, d_t, x^{t+\tau}_{t-\tau})\ dF_{t,T}(y_t|y_0, d^{t-1}, x^T) \\ & \qquad t = 1, \ldots, T-\tau. \end{aligned}$$

As in White and Kennedy (2008), we adopt the convention that covariate values for negative time indexes ($t = -\tau, ..., -1$) are observable. Thus, for example, $X^T \equiv (X_{-\tau}, ..., X_T)$. Assumption A.5 permits us to write x^{t-1} in the integral in Proposition 4.1(a) in place of the term x^t that would otherwise appear using White and Kennedy's argument.

Our next result exploits these recursions to provide a basis for practical tests of various kinds of Granger non-causality.

Proposition 4.2

(a) Suppose Assumptions A.1(a′), A.2(a), A.4(a), and A.5 hold. Then $Y_{t+1} \perp D^t \mid Y_0, X^t$ *for all* $t \in \mathbb{N}$ *if and only if*

$$Y_{t+1} \perp D_t \mid Y_t, X^t_{t-\tau} \quad \textit{for all } t \in \mathbb{N}.$$

(b) Suppose Assumptions A.1(a′), A.2(b), and A.4(b) hold. Then for the given T, $Y_{t+1} \perp D^t \mid Y_0, X^T$ *for all integers* $t < T$ *if and only if for the given* T

$$Y_{t+1} \perp D_t \mid Y_t, X^{t+\tau}_{t-\tau} \quad \textit{for all integers } t \leq T-\tau.$$

It follows that one can test weak G non-causality when A.1(a′), A.2(a), A.4(a), and A.5 hold by testing the equivalent condition $Y_{t+1} \perp D_t \mid Y_t, X^t_{t-\tau}$. Similarly, one can test retrospective weak G non-causality when A.1(a′), A.2(b), and A.4 (b) hold by testing the equivalent condition $Y_{t+1} \perp D_t \mid Y_t, X^{t+\tau}_{t-\tau}$.

These tests are straightforward to perform, as they do not involve entire histories of any of the variables. Other notable features of these conditions are: (i) they explicitly specify which conditioning variables to include, namely a near history of the covariates; (ii) the retrospective version includes leads as well as lags of the covariates.

There are numerous methods for testing the conditional independence identified in Proposition 4.2. For example, one can apply nonparametric methods of Su and White (2007a, 2007b, 2008). Nevertheless, because of the typically modest number of time series observations available relative to the number of relevant observable variables, parametric methods for testing conditional independence will usually be more practical. In the applications of Section 6, we illustrate how parametric methods similar to those proposed in section 5 of White (2006a) can be exploited to obtain tests with useful power against a wide range of departures from conditional independence.

5 Testing (Retrospective) Conditional Exogeneity

Our main results in Section 3 crucially involve either conditional exogeneity or retrospective conditional exogeneity. In this section we provide some discussion about how to choose covariates to help ensure these conditions, and we provide results supporting tests.

For concreteness, consider conditional exogeneity, the property that $D^t \perp U^t \mid Y_0, X^t$. As explained in White (2006a), covariates X_t should include variables that are useful for predicting either D^t or U^t, as the better a predictor X_t is, the less useful D^t will be as a predictor for U^t and vice-versa. By applying Reichenbach's (1956) principle of common cause (if $\mathcal{X}$ and $\mathcal{Y}$ are correlated, then either $\mathcal{X}$ causes $\mathcal{Y}$, $\mathcal{Y}$ causes $\mathcal{X}$, or there is an underlying common cause, say $\mathcal{Z}$), White (2006a) reasons that X_t should include: (1) observable causes $\tilde{V}_t$ and $\tilde{Z}_t$ of Y_{t+1}; (2) observable causes $\tilde{V}_t, W_t$, and $\tilde{Z}_t$ of D_{t+1}; and (3) observable responses W_t to unobservable causes U_t of Y_{t+1} and/or D_{t+1}. In the case of retrospective conditional exogeneity, the latter class of variables justifies inclusion of leads of $\tilde{V}_t, W_t$, and $\tilde{Z}_t$.

Although economic theory can play a key role in justifying conditional exogeneity, it is desirable to have straightforward statistical methods for empirically testing whether a given X_t delivers (retrospective) conditional exogeneity. White (2006a) and White and Chalak (2007c) provide a variety of tests for conditional exogeneity. White (2006a) considers time-series conditional exogeneity tests where D_t is a binary variable. White and Chalak (2007c) provide tests for cross-section data. Here we take a similar approach appropriate for our general dynamic structure.

5.1 Testing conditional exogeneity

A first challenge in testing conditional exogeneity is that U_t is unobservable. A result of White and Chalak (2007c) permits us to overcome this by delivering a consequence of conditional exogeneity that involves only observables, provided that one can observe suitable additional proxies. Assumption A.6(a) describes these additional variables.

Assumption A.6 (a) There exists an observable $k_{\tilde{w}} \times 1$ random vector $\tilde{W}_0$, with $k_{\tilde{w}}$ finite and positive, and an unobservable sequence of $k_{\ddot{u}} \times 1$ random vectors $\{\ddot{U}_t\}$, with $k_{\ddot{u}}$ a countably valued integer, such that (i) $\{\tilde{W}_t\}$ is generated as

$$\tilde{W}_{t+1} \overset{c}{=} b_{3,t+1}(\tilde{W}^t, X^t, U^t, \ddot{U}^t), \quad t = 0, 1, ...,$$

where $b_{3,t+1}$ is an unknown measurable function of dimension $k_{\tilde{w}}$; and (ii)

$$D^t \perp (\ddot{U}^t, \tilde{W}_0) \mid Y_0, U^t, X^t, \quad t = 1, 2, ... \ .$$

A leading example is the case in which $\tilde{W}_{t+1}$ is an error-laden version of U_t, e.g., $\tilde{W}_{t+1} = w(U_t) + \ddot{U}_t$. When $\ddot{U}_t$ is a measurement error, Assumption A.6(a.ii) is quite plausible.

We also impose an assumption on the initial values Y_0 and $\tilde{W}_0$ that permits us to remove Y_0 from the conditioning information set.

Assumption A.7 (a) $(\tilde{W}_{t+1}, \tilde{W}_t) \perp (\tilde{W}_0, Y_0) \mid X^t$ for all $t = 1, 2, ...$.

Assumption A.7(a) imposes a plausible restriction on the usefulness of the initial values of $(\tilde{W}_0, Y_0)$ for predicting $\tilde{W}_t$ and $\tilde{W}_{t+1}$, given X^t.

Tests for conditional exogeneity based solely on observables follow from our next result.

Proposition 5.1 *(a) Suppose Assumptions A.1(a), A.6(a), and A.7(a) hold. Then* $D^t \perp U^t \mid Y_0, X^t$ *for all* $t \in \mathbb{N}$ *implies* $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, X^t$ *for all* $t \in \mathbb{N}$.

As in the case of weak G non-causality, it is inconvenient to attempt to test conditions that involve entire data histories (here, D^t and X^t). Nevertheless, an assumption similar to A.4(a) delivers results similar to those of Proposition 4.2(a), supporting convenient tests for conditional exogeneity that involve only finite data histories. For this we let $dF_{t+1,t}(\ \tilde{w}_{t+1} \mid \tilde{w}_t, d_t, x^t)$ define the conditional density of $\tilde{W}_{t+1}$ given $\tilde{W}_t = \tilde{w}_t, D_t = d_t$, and $X^t = x^t$, and we let $dF_{-\tau}(\ \tilde{w}_{t+1} \mid \tilde{w}_t, d_t, x_{t-\tau}^t)$ define the conditional density of $\tilde{W}_{t+1}$ given $\tilde{W}_t = \tilde{w}_t, D_t = d_t$, and $X_{t-\tau}^t = x_{t-\tau}^t$.

Assumption A.8 (a) There exists a finite non-negative integer τ such that for all $t \in \mathbb{N}$, and for all argument values

$$dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x^t) = dF_{-\tau}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x_{t-\tau}^t).$$

We also impose an assumption analogous to A.5.

Assumption A.9 For all $t \in \mathbb{N}^+, \tilde{W}_t \perp X_t \mid \tilde{W}_0, X^{t-1}$.

A.9 is plausible, as X_t does not enter the structural equation for $\tilde{W}_t$.

The next result supports conditional exogeneity tests that involve only a finite number of lags of X_t.

Proposition 5.2 *(a) Suppose Assumptions A.1(a), A.6(a), A.7(a), A.8(a), and A.9 hold. Then* $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, X^t$ *for all* $t \in \mathbb{N}$ *if and only if* $\tilde{W}_{t+1} \perp D_t \mid \tilde{W}_t, X_{t-\tau}^t$ *for all* $t \in \mathbb{N}$.

Thus, rejecting $\tilde{W}_{t+1} \perp D_t \mid \tilde{W}_t, X_{t-\tau}^t$ implies rejecting conditional exogeneity, $D^t \perp U^t \mid Y_0, X^t$. Just as for (weak) G non-causality tests, one may apply either parametric or non-parametric tests of the conditional independence hypothesis $\tilde{W}_{t+1} \perp D_t \mid \tilde{W}_t, X_{t-\tau}^t$.

5.2 Testing retrospective conditional exogeneity

Results for retrospective conditional exogeneity analogous to those of the previous subsection hold under analogous conditions. The notation in Assumption A.8(b) is parallel to that previously defined.

Assumption A.6 (b) Assumption A.6(a.i) holds; and (ii) for given finite integer T,

$$D^t \perp (\ddot{U}^t, \tilde{W}_0) \mid Y_0, U^t, X^T, \quad t = 1, 2, ..., T.$$

Assumption A.7 (b) $(\tilde{W}_{t+1}, \tilde{W}_t) \perp (\tilde{W}_0, Y_0) \mid X^T$ for all $t = 1, 2, ..., T-1$.

Proposition 5.1 *(b) Suppose Assumptions A.1(a), A.6(b), and A.7(b) hold. Then* $D^t \perp U^t \mid Y_0, X^T$ *for all* $t = 0, 1, ..., T$ *implies* $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, X^T$ *for all* $t = 0, 1, ..., T-1$.

Assumption A.8 (b) There exists a finite non-negative integer τ such that for given T and all integers $t \leq T - \tau$, and for all argument values

$$dF_{t+1,T}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x^T) = dF_{\tau}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x_{t-\tau}^{t+\tau}).$$

Proposition 5.2 *(b) Suppose Assumptions A.1(a), A.6(b), A.7(b), and A.8(b) hold. Then for given* T*,* $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, X^T$ *for all integers* $t < T$ *if and only if* $\tilde{W}_{t+1} \perp D_t \mid \tilde{W}_t, X_{t-\tau}^{t+\tau}$ *for all integers* $t \leq T - \tau$*.*

Observe that we do not require A.9 here. Parallel to the previous case, rejecting $\tilde{W}_{t+1} \perp D_t \mid \tilde{W}_t, X_{t-\tau}^{t+\tau}$ implies rejecting retrospective conditional exogeneity, $D^t \perp U^t \mid Y_0, X^T$.

5.3 A pure test of structural non-causality

Our results of Section 3 say that if we reject (weak) G non-causality, then we must reject either structural non-causality or conditional exogeneity (or both). The results just given provide a way to test conditional exogeneity. If we find that we reject G non-causality but not conditional exogeneity, then we have evidence against structural non-causality. We thus propose a simple formal test of structural non-causality:

> Reject structural non-causality if the (retrospective) (weak) G non-causality test *rejects* and the (retrospective) conditional exogeneity test *fails to reject.*

Our next result provides easy bounds on the level and power of this test.

Proposition 5.3 *Suppose the significance levels of the (retrospective) conditional exogeneity test, the (retrospective) (weak)* G *non-causality test, and the structural non-causality test are* α_1, α_2*, and* α *respectively. Suppose the powers of the (retrospective) conditional exogeneity test, the (retrospective) (weak)* G *non-causality test, and the structural non-causality test are* π_1, π_2*, and* π *respectively. Then*

$$\max\{0, \min\{(\alpha_2 - \alpha_1), (\pi_2 - \pi_1), (\alpha_2 - \pi_1)\}\} \leq$$
$$\alpha \leq \max\{\min\{1-\alpha_1, \alpha_2\}, \min\{1-\pi_1, \pi_2\}, \min\{1-\pi_1, \alpha_2\}\} \quad \textit{and}$$

$$\pi_2 - \alpha_1 \leq \pi \leq \min\{1 - \alpha_1, \pi_2\}.$$

The intuition here is that the level of the structural non-causality test decreases with the level of the G non-causality test and the power of the conditional exogeneity test. The lower bound for the level is zero when π_1 is sufficiently large. For the power, the intuition is that the power of the structural non-causality test increases with the power of the G non-causality test and decreases with the level of the conditional exogeneity test.

In applications, tests are usually conducted using asymptotic critical values. The exact level and power of a given test is then unknown, but it nevertheless converges to a known limit. When the conditional exogeneity and G non-causality tests are consistent, as can often be arranged, we have the following properties for the asymptotic level and power of our procedure. (Limits are taken as $T \to \infty$.)

Proposition 5.4 *Suppose that for $T = 1, 2, ...$ the significance levels of the (retrospective) conditional exogeneity test, the (retrospective) (weak) G non-causality test, and the structural non-causality test are α_{1T}, α_{2T}, and α_T respectively, and that $\alpha_{1T} \to \alpha_1$ and $\alpha_{2T} \to \alpha_2$. Suppose the powers of the (retrospective) conditional exogeneity test, the (retrospective) (weak) G non-causality test, and the structural non-causality test are π_{1T}, π_{2T}, and π_T respectively, and that $\pi_{1T} \to 1$ and $\pi_{2T} \to 1$. Then*

$$0 \leq \liminf \alpha_T \leq \limsup \alpha_T \leq \min\{1 - \alpha_1, \alpha_2\} \quad \textit{and} \quad \pi_T \to 1 - \alpha_1.$$

Typically, we can achieve $\alpha_1 = 0$ and $\alpha_2 = 0$ for a consistent test by suitable choice of an increasing sequence of critical values. In this case we also have $\alpha_T \to 0$ and $\pi_T \to 1$. We note that conditional exogeneity tests may or may not be consistent against every possible alternative. When the DGP corresponds to an alternative for which unit power is not achieved, a weaker bound on the level and power of structural non-causality test still holds by Proposition 5.3. This suggests exercising care to design the conditional exogeneity test to be consistent against particularly important or plausible alternatives.

6 Illustrative Applications

In this section, we illustrate our methods by studying structural causality in industrial pricing, macroeconomics, and finance. First, we investigate structural causality from crude oil prices to gasoline prices. Second, we examine structural causality from monetary policy to industrial production. Third, we investigate structural causality from expected macroeconomic announcements to stock returns.

6.1 Test Implementation

To test (retrospective) weak G non-causality and (retrospective) conditional exogeneity, we require tests for conditional independence. Non-parametric tests for conditional independence consistent against arbitrary alternatives are readily available (e.g., Linton and

Gozalo, 1997; Fernandes and Flores, 2001; Delgado and Gonzalez-Manteiga, 2001; and Su and White, 2007a, 2007b, 2008), but given the large number of covariates in most applications, these are often not practical. Here we apply parametric tests designed to have power against a relatively broad spectrum of alternatives. Specifically, we apply both traditional linear parametric tests and new, more flexible parametric tests based on the "QuickNet" procedures introduced by White (2006b).

6.1.1 Testing conditional mean independence with linear conditional expectations

Traditional tests of G non-causality are based on the fact that rejection of conditional mean independence implies rejection of conditional independence. Tests of conditional mean independence are typically implemented by assuming the linearity of the conditional expectation. Accordingly, to test $\mathcal{Y} \perp \mathcal{D} \mid \mathcal{S}$, we first suppose that

$$E(\mathcal{Y} \mid \mathcal{D}, \mathcal{S}) = \alpha + \mathcal{D}'\beta_0 + \mathcal{S}'\beta_1.$$

Under the null that $\mathcal{Y} \perp \mathcal{D} \mid \mathcal{S}$, we have $\beta_0 = 0$. Thus, we estimate the coefficients of the following regression equation and test $\beta_0 = 0$:

$$\mathcal{Y} = \alpha + \mathcal{D}'\beta_0 + \mathcal{S}'\beta_1 + \varepsilon. \qquad \text{(CI Test Regression 1)}$$

If we reject $\beta_0 = 0$, then we also reject $\mathcal{Y} \perp \mathcal{D} \mid \mathcal{S}$. On the other hand, if we fail to reject $\beta_0 = 0$, this does not necessarily imply a failure to reject $\mathcal{Y} \perp \mathcal{D} \mid \mathcal{S}$, as conditional independence can easily fail even if conditional mean independence holds.

6.1.2 Testing conditional mean independence with flexible conditional expectations

To achieve power against a wider range of alternatives than the traditional linear method just specified, we can use a more flexible function to test conditional mean independence. In particular, we consider a specification exploited in White's (2006b, p.476) QuickNet procedure:

$$E(\mathcal{Y} \mid \mathcal{D}, \mathcal{S}) = a + \mathcal{D}'\beta_0 + \mathcal{S}'\beta_1 + \sum_{j=1}^{q} \psi(\mathcal{S}'\gamma_j)\beta_{j+1},$$

where ψ is a given activation function belonging to the class of generically comprehensively revealing (GCR) functions. Here we let ψ be the logistic cdf $\psi(z) = 1/(1 + \exp(-z))$ or the ridgelet function $\psi(z) = (-z^5 + 10z^3 - 15z)\exp(-.5z^2)$ (see, for example, Candès, 1999). We call $\psi(\mathcal{S}'\gamma_j)$ the "activation" of "hidden unit" j. The integer q lies between 1 and $\bar{q}$, the maximum number of hidden units. We choose γ_j according to the algorithm proposed in White (2006b, p.477). Thus, we estimate the coefficients of the following regression equation and test $\beta_0 = 0$:

$$\mathcal{Y} = a + \mathcal{D}'\beta_0 + \mathcal{S}'\beta_1 + \sum_{j=1}^{q} \psi(\mathcal{S}'\gamma_j)\beta_{j+1} + \varepsilon. \qquad \text{(CI Test Regression 2)}$$

6.1.3 Testing conditional independence using non-linear transformations

Third, to gain power against alternatives for which conditional mean independence holds but conditional independence fails, we consider tests based on transformations of $\mathcal{Y}$ and $\mathcal{D}$, as $\mathcal{Y} \perp \mathcal{D} \mid \mathcal{S}$ implies that $\psi_y(\mathcal{Y}) \perp \psi_d(\mathcal{D}) \mid \mathcal{S}$ for all measurable functions ψ_y and ψ_d. Accordingly, we consider specifications of the form

$$E(\psi_y(\mathcal{Y}) \mid \psi_d(\mathcal{D}), \mathcal{S}) = a + \psi_d(\mathcal{D})'\beta_0 + \mathcal{S}'\beta_1 + \sum_{j=1}^{q} \psi(\mathcal{S}'\gamma_j)\beta_{j+1}.$$

We let ψ_y and ψ_d be the logistic cdf or ridgelet functions. The choices of γ, q, and ψ are the same as in the previous case. Thus, we estimate the coefficients of the following regression equation and test $\beta_0 = 0$:

$$\psi_y(\mathcal{Y}) = a + \psi_d(\mathcal{D})'\beta_0 + \mathcal{S}'\beta_1 + \sum_{j=1}^{q} \psi(\mathcal{S}'\gamma_j)\beta_{j+1} + \varepsilon. \qquad \text{(CI Test Regression 3)}$$

6.2 Crude oil and gasoline prices

White and Kennedy (2008) apply their methods for estimating retrospective causal effects to study the impact of crude oil prices on gasoline prices. Our methods here are fully applicable to this setting. We let the response of interest, Y_t, be the natural logarithm of the spot price for US Gulf Coast conventional gasoline; our cause of interest, D_t, is the natural logarithm of the Cushing OK WTI spot crude oil price. We let U_t represent all unobservable causes of gasoline prices, so that $\tilde{V}_t$ and $\tilde{Z}_t$ have zero dimension, as in White and Kennedy (2008).

Considering that crude oil prices should be quickly reflected in gasoline prices (e.g., as found by Borenstein, Cameron, and Gilbert, 1997) and that our data frequency is monthly, a relatively long interval, here we investigate the structural causality of D^t on contemporaneous Y_t. Thus, we take the structure of interest to be

$$Y_t \overset{c}{=} r_t(Y_0, D^t, U^t), \quad t = 0, 1, ..., T.$$

We view this as involving "contemporaneous" rather than "instantaneous" causation. We distinguish these notions as follows. For contemporaneous causation, the cause precedes the response, but the causal interval is less than the observational interval. Thus, the causal response occurs within the observed interval, justifying the appearance of the contemporaneous value of a potential cause in the structural relation. On the other hand, instantaneous causation violates the principle that causes precede effects. We rule this out.

We do not explicitly treat contemporaneous causation in our theoretical analysis above, but all the relevant results hold by replacing Y_{t+1} with Y_t on the left-hand side of the implicit dynamic response function, and similarly in the other implicit dynamic relations. This structure is that explicitly analyzed by White and Kennedy (2008).

Let W_t be proxies for the unobservable causes U_t. Similar to White and Kennedy (2008), we let W_t include (1) the natural logarithm of Texas Initial and Continuing Unemployment Claims (taken from State Weekly Claims for Unemployment Insurance Data, Not Seasonally Adjusted); (2) Houston temperature; (3) a winter dummy for January, February, and March; (4) a summer dummy for June, July, and August; (5) the natural logarithm of U.S. Bureau of Labor Statistics Electricity price index; (6) the 10-Year Treasury Note Constant Maturity Rate; (7) the 3-Month T-Bill Secondary Market Rate; and (8) the Index of the Foreign Exchange Value of the Dollar. Here, the covariates are $X_t = W_t$. Our sample period covers from January 1987 through December 1997, a total 132 observations.

Over our sample period, both crude oil and gasoline prices are relatively stable. Specifically, the augmented Dickey-Fuller tests for stationarity of Y_t and D_t reject the null hypothesis that these series contain a unit root. Nevertheless, we cannot reject the hypothesis that there is a unit root for certain of the covariates (the 10-year Treasury Note rate, the 3-month T-Bill rate, the natural logarithm of the electricity price index, and the index of the foreign exchange value of the dollar). We enter these covariates in first differences.

There are some differences between our use of data and that of White and Kennedy (2008): (1) The sample period is different; (2) The natural gas price index does not appear in W_t; instead we include this in $\tilde{W}_t$ below; (3) We use the Federal Reserve's Index of the Foreign Exchange Value of the Dollar instead of the Yen-US dollar and British pound-US dollar exchange rates to avoid multicollinearity (see http://www.federalreserve.gov/releases/h10/summary/); and (4) Here, we consider a stationary process with contemporaneous causation, whereas White and Kennedy's (2008) empirical analysis involves a cointegrated relation with a one period lag.

To test structural non-causality, we apply the procedure of Section 5.3. Specifically, we perform tests of the key retrospective conditional exogeneity assumption, $D^t \perp U^t | Y_0, X^T$, together with tests for retrospective weak G non-causality.

To test retrospective conditional exogeneity, we let $\tilde{W}_t$ be the natural logarithm of the U.S. Bureau of Labor Statistics Natural Gas Price Index. The key requirement for $\tilde{W}_t$ is A.6(b). Allowing contemporaneous causation, we state A.6 (b.i) as

$$\tilde{W}_t \overset{c}{=} b_{3,t}(\tilde{W}^{t-1}, X^t, U^t, \ddot{U}^t),$$

where $\ddot{U}_t$ represents additional unobservable drivers of $\tilde{W}_t$ other than U_t. This is a plausible assumption, as we may view $\tilde{W}_t$ as an error-laden measure of the natural gas prices that actually drive oil prices, a component of U_t. The measurement error is then $\ddot{U}_t$. Assumption A.6 (b.ii) is the condition that

$$D^t \perp (\ddot{U}^t, \tilde{W}_0) \mid Y_0, U^t, X^T.$$

This condition means that conditioning on all the unobservable causes of gasoline prices and the covariates, crude oil prices are independent of the measurement error history $\ddot{U}^t$ and the initial value of the natural gas price index, $\tilde{W}_0$. This is also reasonably plausible.

With suitable memory and conditional stationarity assumptions as specified in Section 5, we test retrospective conditional exogeneity by testing $D_t \perp \tilde{W}_t \mid \tilde{W}_{t-1}, X_{t-\tau}^{t+\tau}$. We implement the tests in the three forms (CI Test Regressions 1-3) described above. Specifically, we let $\mathcal{Y} = \tilde{W}_t$, $\mathcal{D} = D_t$, and $\mathcal{S} = (\tilde{W}_{t-1}, X_{t-\tau}^{t+\tau})$, and we test $\beta_0 = 0$. Given the relatively small sample size, we consider only $\tau \leq 5$.

We report our results in Tables 1-3. For CI Test Regression 1, we cannot reject $\beta_0 = 0$ for all $\tau = 1, 2, ..., 5$ at the 5% significance level. For CI Test Regression 2, we take $\bar{q} \leq 5$. Letting ψ be the ridgelet function, we again cannot reject $\beta_0 = 0$ for all τ and q at the 5% significance level. If we let ψ function be the logistic cdf, we cannot reject $\beta_0 = 0$ for most values of τ and q. Exceptions are for $\tau = 2$ and $q = 2, 3, 4$ or 5; however, the Bonferroni-Hochberg p-value bounds for the rows and columns and for the table as a whole do not support rejection. For CI Test Regression 3, we again cannot reject $\beta_0 = 0$ for almost all the choices of τ, q, ψ, $\psi_{\tilde{w}}$, and ψ_d. For brevity, we report results only for the ridgelet case. Taken together, our results suggest that we cannot reject retrospective conditional exogeneity.

For comparison purposes, we also implemented a (non-retrospective) conditional exogeneity test of the hypothesis $D_t \perp \tilde{W}_t \mid \tilde{W}_{t-1}, X_{t-\tau}^{t}$, again using CI Test Regressions 1-3.

The results are quite similar: we again cannot reject conditional exogeneity. Given the similarity of the results, to save space we report only the results for CI Test Regression 1 (see Table 1).

Next, we implement retrospective weak G non-causality tests. With suitable memory and conditional stationarity assumptions as specified in Section 4, we can test retrospective weak G non-causality by testing $Y_t \perp D_t \mid Y_{t-1}, X_{t-\tau}^{t+\tau}$. We implement the tests in the three forms described above. Specifically, we let $\mathcal{Y} = Y_t$, $\mathcal{D} = D_t$, and $\mathcal{S} = (Y_{t-1}, X_{t-\tau}^{t+\tau})$, and we test $\beta_0 = 0$. Tables 4-6 contain the results. As expected, we soundly reject $\beta_0 = 0$ for all three regressions and for all choices of τ, q, ψ, ψ_y, and ψ_d. As a comparison, we also implement (non-retrospective) weak G non-causality tests. The results are similar; again we soundly reject $\beta_0 = 0$ for all the three regressions. To save space, we report these results only in Table 4.

Given the consistency of our tests for retrospective conditional exogeneity, Proposition 5.4 ensures that we can soundly reject the hypothesis of structural non-causality from crude oil prices to contemporaneous gasoline prices. Interestingly, when we replace contemporaneous crude oil prices with lagged crude oil prices, the results become much more equivocal, suggesting that contemporaneous causation plays a central role in this market. For brevity, we do not tabulate these results here.

6.3 Monetary policy and industrial production

Angrist and Kuersteiner (2004) study the causal relationship between the Federal Reserve's monetary policy and output using the data from Romer and Romer (1989). Romer and Romer (1989, 1994) construct a monetary policy shock variable using a narrative approach. They examine the Federal Open Market Committee minutes to identify dates when the Fed took a marked anti-inflationary stance. There are six such dates over the period from 1948 through 1991 (Romer and Romer 1994), the "Romer dates." The total number of observations is 528.

Our methods are applicable to this setting. As in Angrist and Kuersteiner (2004), we let the response of interest (Y_t) be industrial production growth and the cause of interest (D_t) be the Fed's anti-inflationary stance as measured by the Romer dates. The observable ancillary causes ($\tilde{Z}_t$) are unemployment and inflation rates, and U_t represents unobservable causes. W_t and $\tilde{V}_t$ are empty, so here the covariates are $X_t = \tilde{Z}_t$. Industrial production is thus determined as

$$Y_{t+1} \overset{c}{=} r_{t+1}(Y_0, D^t, \tilde{Z}^t, U^t), \quad t = 0, 1, ..., T.$$

The null hypothesis is that monetary policy has no causal effect on the real economy. Under this null, not only does D^t not structurally cause Y_{t+1}, but it also does not structurally cause $\tilde{Z}^t$ or U^t. This system thus satisfies the recursivity required by Assumption A.1.

Significantly, this structure also justifies the key retrospective conditional exogeneity assumption, $D^t \perp U^t \mid X^T$. This condition says that given X^T, U^t cannot help predict D^t, and vice versa. Put another way, given past and future unemployment and inflation rates, the Fed's policy is as good as randomly assigned. Fed decisions usually target future inflation and/or unemployment rates. To the extent that expectations about these targets are driven by past and present values of unemployment and inflation, current and lagged values of X_t should predict D^t well, leaving little role for U^t in predicting D^t. Even when the null is true, so that D^t has no impact on future values of X_t, those future values may be driven by U^t. Leads of X_t may thus be useful in back-casting U^t, leaving little role for D^t in predicting U^t. Both of these features act to ensure $D^t \perp U^t \mid X^T$. Moreover, when the null is false, so that D^t does impact future values of X_t, leads of X_t will act to back-cast D^t, again reducing any role of U^t in predicting D^t.

An important criticism of Romer and Romer's (1989) approach is that it neglects the forward-looking aspects of monetary policy (e.g., Shapiro, 1994; Leeper, 1997). By appropriately conditioning on both leads and lags of covariates, we exploit rather than neglect the forward-looking aspects of monetary policy. This makes the use of our retrospective approach especially appealing.

These structural considerations justify taking retrospective conditional exogeneity to be given here. Testing structural non-causality can now be accomplished by testing retrospective weak G non-causality. With suitable memory and conditional stationarity conditions, as specified in Section 4, we test this by testing $Y_{t+1} \perp D_t \mid Y_t, X_{t-\tau}^{t+\tau}$. We run the three CI test regressions described above; specifically, we let $\mathcal{Y} = Y_{t+1}$, $\mathcal{D} = D_t$, and $\mathcal{S} = (Y_t, X_{t-\tau}^{t+\tau})$ and test $\beta_0 = 0$, letting the maximum τ equal 12. For CI Test Regression 1, we find that we cannot reject $\beta_0 = 0$, no matter how we choose τ. The results are reported in Table 7. For CI Test Regression 2, we let the maximum q equal 9 and we again find that β_0 is insignificant for all τ. Representative results are reported in Table 8. For CI Test Regression 3, again we find that β_0 is insignificant for most choices of τ, q, ψ, ψ_y and ψ_d. Representative results are reported in Table 9.

For comparison, we test weak G non-causality conditioning only on covariate lags, i.e., we test $Y_{t+1} \perp D_t \mid Y_t, X_{t-\tau}^{t}$. For CI Test Regression 1, we find that β_0 is insignificant when $\tau < 4$ and significant when $\tau \geq 4$. These results are reported in Table 7. For CI

Test Regressions 2 and 3, the results are the same as in the retrospective case: we cannot reject $\beta_0 = 0$. We omit tabulating these results to save space.

Before concluding that there is no evidence against the null that monetary policy has no real impacts, we must consider the possibility that the effects of monetary policy take more than one month to be felt. Thus, we relax the single lag assumption of Section 4, and we now permit monetary policy over the past year (D_{t-12}^t) to possibly impact the current growth of industrial production. We continue to assume retrospective conditional exogeneity. Under the null of no structural causality, we then have $Y_{t+1} \perp D_{t-12}^t \mid Y_t, X_{t-\tau}^{t+\tau}$. We again run the three CI test regressions, but now with $\mathcal{Y} = Y_{t+1}$, $\mathcal{D} = D_{t-12}^t$, and $\mathcal{S} = (Y_t, X_{t-\tau}^{t+\tau})$, and we test $\beta_0 = 0$ jointly. For all three CI test regressions, we soundly reject $\beta_0 = 0$. Representative results are reported in Tables 10-12.

For comparison, we also condition only on lags of covariates, i.e., we test $Y_{t+1} \perp D_{t-12}^t \mid Y_t, X_{t-\tau}^t$. Again we reject $\beta_0 = 0$. The results of CI Test Regression 1 are reported in Table 10. The results of Regressions 2 and 3 are the same; we omit these to save space.

Our results suggest that we can soundly reject the hypothesis that monetary policy has no causal effects on the real economy if we permit the effects of monetary policy to gradually filter into the real economy. This is consistent with Romer and Romer's (1989, 1994) conclusion that "these [monetary] policy shifts were followed by large and statistically significant declines in real output relative to its usual behavior. We interpret these results as supporting the view that monetary policy has substantial real effects." On the other hand, our results contrast with Angrist and Kuersteiner's (2004) conclusions; they find that "money-output causality can fairly be described as mixed."

6.4 Stock returns and macroeconomic announcements

Beginning with Chen, Roll, and Ross (1986), many authors have studied the impact of macroeconomic factors on aggregate stock returns (see, e.g., Flannery and Protopapadakis, 2002). In this section, we investigate whether there are causal effects from expected economic announcements to stock returns. This is in part a test of market weak efficiency, because if stock markets are efficient in the weakest sense, then expected returns should not respond to expected economic announcements. On the other hand, other moments of the returns distribution may be affected by expected announcements without violating market weak efficiency.

Although it would also be interesting to examine the causal effects of economic news (unexpected announcements) on stock returns, it is not obvious how one might justify (retrospective) conditional exogeneity for news. We therefore leave an investigation of the effects of news to other work.

We let Y_t, D_t, and $\tilde{V}_t$ denote stock market returns, expected economic announcements, and economic news, respectively; and we let U_t denote unobservable causes. The structural relation is thus

$$Y_{t+1} \stackrel{c}{=} r_{t+1}(Y_0, D^t, \tilde{V}^t, U^t), \quad t = 0, 1, ..., T$$

The sample consists of daily data from January 5, 1995 through October 31, 2006. The daily returns series is that for the value-weighted NYSE-AMEX-NASDAQ market index from the Center for Research on Security Prices (CRSP).

We decompose macroeconomic announcements into economic news and expected changes. For example, let A_t denote a macroeconomic announcement at time t and let E_t denote its expectation. Then $A_t - A_{t-1} = (A_t - E_t) + (E_t - A_{t-1}) = \tilde{V}_t + D_t$; $\tilde{V}_t = A_t - E_t$ then represents news and $D_t = E_t - A_{t-1}$ represents the expected change.

Here we include eight major macroeconomic announcements: (1) real GDP (advanced); (2) core CPI; (3) core PPI; (4) unemployment rate; (5) new home sales; (6) nonfarm payroll employment; (7) consumer confidence; and (8) capacity utilization rate. The expectations of these announcements are gathered from the Money Market Service, which surveys the expectations of professionals and practitioners for those series scheduled to be announced during the following week. These data are widely to represent expectations of macroeconomic variables. To make the expected and unexpected announcements comparable and unit free, we divide each by its standard deviation.

We let W_t represent drivers of D_t as well as responses to unobservable causes. W_t includes (1) the three month T-Bill yield; (2) the term structure premium, measured by the difference between the yield to maturity of the ten-year bond and the three-month T-Bill; (3) the corporate bond premium, measured by the difference in the yield to maturity between Moody's BAA and AAA corporate bond indexes; (4) the daily change of the Index of the Foreign Exchange Value of the Dollar; (5) the daily change of the crude oil price. The first four variables are computed using data from the U.S. Federal Reserve and the fifth from the Energy Information Administration. We view these variables as representing macroeconomic fundamentals. The covariates are $X_t = (W_t, \tilde{V}_t)$.

We use the augmented Dickey-Fuller test to test the stationarity of Y_t and D_t, and we reject the hypothesis that there is a unit root for both of them. Nevertheless, for some covariates (the three month T-Bill yield, the corporate bond premium, and the term structure premium), we cannot reject the unit root hypothesis; in these cases, we use first differences as covariates.

Again, we take retrospective conditional exogeneity, $D^t \perp U^t \mid X^T$, as given. This is plausible, as the covariates include leads and lags of macroeconomic news and other macroeconomic fundamentals. Given these, one would not expect unobservable causes of stock returns to predict investors' expectations of changes in macroeconomic announcements.

With suitable memory and conditional stationarity conditions as specified in Section 4, we can test structural non-causality by testing retrospective weak G non-causality, based on $Y_{t+1} \perp D_t \mid Y_t, X_{t-\tau}^{t+\tau}$. We again run the three forms of the CI test regressions, letting $\mathcal{D} = D_t, \mathcal{Y} = Y_{t+1}$, and $\mathcal{S} = (Y_t, X_{t-\tau}^{t+\tau})$, and we test $\beta_0 = 0$ jointly.

For CI Test Regression 1, we let the maximum τ equal 7. We find that for all τ, we cannot reject $\beta_0 = 0$. Table 13 contains the results. For CI Test Regression 2, we let the maximum q equal 9. Again, we do not reject $\beta_0 = 0$, as reported in Table 14. These results are consistent with weak market efficiency.

For CI Test Regression 3, when ψ_y and ψ_d are logistic functions, we again do not reject $\beta_0 = 0$ for all τ and q. Nevertheless, for the ridgelet function case, we do reject $\beta_0 = 0$ for all τ and q, as reported in Table 15. Thus, expected macroeconomic announcements do appear to have a structural impact on stock returns, but not on mean returns, consistent with market weak efficiency. We also note that this pattern of results is consistent with our maintained assumption of retrospective conditional exogeneity. If this assumption did not hold, we would expect the test for retrospective weak G non-causality to reject across the board.

For comparison, we perform the same tests, conditioning on lags only, testing $Y_{t+1} \perp D_t \mid Y_t, X_{t-\tau}^{t}$. The results exhibit the identical pattern. To save space, we report only the results of CI Test Regression 1 in Table 13.

7 Summary and Concluding Remarks

In this paper, we specify a general nonseparable recursive dynamic structural system and give a natural definition of structural causality for such systems. Building on classical notions of G non-causality, we introduce interesting and natural extensions, namely weak G non-causality and retrospective weak G non-causality. We show that structural non-causality and (retrospective) conditional exogeneity imply (retrospective) (weak) G non-causality. We strengthen structural causality to notions of (retrospective) strong causality and show that (retrospective) strong causality implies (retrospective) weak $G-$causality.

We provide practical conditions and straightforward methods for testing (retrospective) weak G non-causality, (retrospective) conditional exogeneity, and structural non-causality. Finally, we apply our methods to explore structural causality in industrial pricing, macroeconomics, and finance.

There are many interesting topics for further research lying beyond the scope of this paper. First, we consider only recursive structures here. It is of definite interest to investigate the relations between structural non-causality and G non-causality in non-recursive systems. Second, it is of interest to explicitly incorporate cointegration into our framework. Although our framework admits cointegrated systems, explicit examination of the relationships between structural non-causality, conditional exogeneity, cointegration, and G non-causality is of special importance. Finally, it is of interest to study the behavior of $\sqrt{n}-$consistent *nonparametric* conditional independence tests implementing (retrospective) conditional exogeneity and (retrospective) weak G non-causality tests when there is a relatively large number of covariates.

Table 1
Crude Oil and Gasoline Prices
(Retrospective) conditional exogeneity test: CI Test Regression 1

τ	conditioning on both leads and lags	conditioning on lags only	row BH
0	0.373	0.373	0.373
1	0.272	0.430	0.430
2	0.060	0.272	0.120
3	0.142	0.201	0.201
4	0.283	0.210	0.283
5	0.171	0.272	0.272
col BH			
	0.359	0.430	0.430

Notes: Numbers in the main entries are individual p-values. BH denotes Bonferroni-Hochberg adjusted p-values. The final diagonal element is the BH p-value for the table as a whole. We use Newey-West (1987) standard errors to compute individual p-values.

Table 2
Crude Oil and Gasoline Prices
Retrospective conditional exogeneity test: CI Test Regression 2 (ψ: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	row BH
0	0.471	0.345	0.298	0.305	0.362	0.471
1	0.201	0.275	0.235	0.212	0.148	0.275
2	0.077	0.063	0.185	0.119	0.094	0.185
3	0.145	0.162	0.157	0.119	0.080	0.162
4	0.396	0.203	0.100	0.108	0.076	0.324
5	0.094	0.137	0.114	0.261	0.210	0.261
col BH						
	0.462	0.345	0.298	0.305	0.362	0.471

Notes: See notes to Table 1.

Table 3
Crude Oil and Gasoline Prices
Retrospective conditional exogeneity test: CI Test Regression 3
($\psi, \psi_{\tilde{w}}, \psi_d$: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	row BH
0	0.846	0.930	0.628	0.916	0.942	0.942
1	0.722	0.898	0.928	0.896	0.806	0.928
2	0.455	0.211	0.205	0.242	0.147	0.455
3	0.991	0.827	0.684	0.303	0.191	0.955
4	0.797	0.791	0.563	0.522	0.574	0.797
5	0.675	0.608	0.752	0.757	0.839	0.839
col BH						
	0.991	0.930	0.928	0.916	0.882	0.991

Notes: See notes to Table 1.

Table 4
Crude Oil and Gasoline Prices
(Retrospective) weak G non-causality test: CI Test Regression 1

τ	conditioning on both leads and lags	conditioning on lags only	row BH
0	0.000	0.000	0.000
1	0.000	0.000	0.000
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.000	0.000	0.000
5	0.000	0.000	0.000
col BH			
	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 5
Crude Oil and Gasoline Prices
Retrospective weak G non-causality test: CI Test Regression 2 (ψ: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	row BH
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
col BH						
	0.000	0.000	0.000	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 6
Crude Oil and Gasoline Prices
Retrospective weak G non-causality test: CI Test Regression 3
(ψ, ψ_y, ψ_d: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	row BH
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
col BH						
	0.000	0.000	0.000	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 7
Monetary Policy (D_t) and Industrial Production
(Retrospective) weak G non-causality test: CI Test Regression 1

τ	conditioning on both leads and lags	conditioning on lags only	row BH
0	0.143	0.143	0.143
1	0.120	0.125	0.125
2	0.183	0.103	0.183
3	0.210	0.053	0.106
4	0.182	0.043	0.086
5	0.196	0.034	0.068
6	0.213	0.028	0.056
7	0.252	0.028	0.056
8	0.254	0.020	0.040
9	0.304	0.026	0.052
10	0.256	0.019	0.038
11	0.351	0.021	0.042
12	0.321	0.032	0.064
col BH			
	0.351	0.143	0.351

Notes: See notes to Table 1.

Table 8
Monetary Policy (D_t) and Industrial Production
Retrospective weak G non-causality test: CI Test Regression 3 (ψ: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.157	0.161	0.071	0.075	0.066	0.082	0.212	0.212	0.204	0.212
1	0.119	0.121	0.114	0.127	0.125	0.136	0.165	0.141	0.099	0.165
2	0.160	0.197	0.215	0.295	0.385	0.376	0.380	0.364	0.362	0.385
3	0.160	0.183	0.202	0.186	0.142	0.064	0.062	0.087	0.081	0.202
4	0.174	0.185	0.216	0.190	0.190	0.196	0.170	0.176	0.147	0.216
5	0.210	0.180	0.178	0.187	0.198	0.206	0.196	0.190	0.163	0.210
6	0.197	0.185	0.181	0.189	0.202	0.223	0.212	0.212	0.193	0.223
7	0.244	0.241	0.226	0.235	0.230	0.154	0.112	0.098	0.138	0.244
8	0.265	0.250	0.318	0.348	0.392	0.339	0.313	0.320	0.301	0.392
9	0.263	0.369	0.443	0.440	0.439	0.391	0.399	0.414	0.438	0.443
10	0.278	0.269	0.279	0.290	0.313	0.261	0.214	0.260	0.333	0.333
11	0.320	0.304	0.283	0.266	0.277	0.209	0.211	0.148	0.143	0.320
12	0.282	0.270	0.308	0.292	0.299	0.310	0.352	0.378	0.348	0.378
col BH										
	0.320	0.369	0.443	0.440	0.439	0.391	0.399	0.414	0.438	0.443

Notes: See notes to Table 1.

Table 9
Monetary Policy (D_t) and Industrial Production
Retrospective weak G non-causality test: CI Test Regression 3
(ψ, ψ_y, ψ_d: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.178	0.172	0.113	0.097	0.054	0.070	0.075	0.052	0.048	0.178
1	0.317	0.228	0.193	0.180	0.167	0.164	0.119	0.117	0.145	0.317
2	0.165	0.112	0.173	0.123	0.124	0.128	0.081	0.076	0.060	0.173
3	0.150	0.165	0.045	0.042	0.016	0.021	0.016	0.013	0.008	0.072
4	0.197	0.245	0.230	0.230	0.269	0.128	0.121	0.126	0.133	0.269
5	0.212	0.107	0.055	0.074	0.052	0.077	0.082	0.095	0.051	0.212
6	0.369	0.360	0.376	0.199	0.175	0.169	0.142	0.162	0.147	0.376
7	0.176	0.297	0.257	0.150	0.120	0.088	0.096	0.066	0.046	0.297
8	0.215	0.154	0.184	0.203	0.141	0.141	0.126	0.088	0.103	0.215
9	0.102	0.266	0.292	0.240	0.247	0.272	0.276	0.284	0.257	0.292
10	0.479	0.473	0.464	0.491	0.423	0.265	0.348	0.364	0.304	0.491
11	0.250	0.320	0.293	0.219	0.182	0.226	0.170	0.173	0.146	0.320
12	0.300	0.263	0.257	0.303	0.199	0.196	0.258	0.296	0.243	0.303
col BH										
	0.479	0.473	0.464	0.491	0.208	0.272	0.208	0.169	0.104	0.491

Notes: See notes to Table 1.

Table 10
Monetary Policy (D_{t-12}^{t}) and Industrial Production
(Retrospective) weak G non-causality test: CI Test Regression 1

τ	conditioning on both leads and lags	conditioning on lags only	row BH
0	0.000	0.000	0.000
1	0.000	0.000	0.000
2	0.000	0.000	0.000
3	0.000	0.000	0.000
4	0.000	0.000	0.000
5	0.000	0.000	0.000
7	0.000	0.000	0.000
8	0.000	0.000	0.000
9	0.000	0.000	0.000
10	0.000	0.000	0.000
11	0.000	0.000	0.000
12	0.000	0.000	0.000
col BH			
	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 11
Monetary Policy (D_{t-12}^{t}) and Industrial Production
Retrospective weak G non-causality test: CI Test Regression 2 (ψ: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
col BH										
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 12
Monetary Policy (D_{t-12}^{t}) and Industrial Production
Retrospective weak G non-causality test: CI Test Regression 3
(ψ, ψ_y, ψ_d: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
col BH										
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: See notes to Table 1.

Table 13
Stock Returns and Macroeconomic Announcements
(Retrospective) weak G non-causality test: CI Test Regression 1

τ	conditioning on both leads and lags	conditioning on lags only	row BH
0	0.818	0.818	0.818
1	0.814	0.833	0.833
2	0.673	0.782	0.782
3	0.658	0.756	0.756
4	0.684	0.750	0.750
5	0.613	0.724	0.724
6	0.514	0.641	0.641
7	0.510	0.663	0.663
col BH			
	0.818	0.833	0.833

Notes: See notes to Table 1.

Table 14
Stock Returns and Macroeconomic Announcements
Retrospective weak G non-causality test: CI Test Regression 2 (ψ: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.804	0.793	0.781	0.790	0.765	0.765	0.744	0.724	0.697	0.804
1	0.782	0.810	0.828	0.848	0.834	0.823	0.824	0.814	0.800	0.848
2	0.663	0.646	0.599	0.608	0.592	0.608	0.644	0.656	0.648	0.663
3	0.653	0.635	0.643	0.621	0.636	0.632	0.608	0.596	0.601	0.653
4	0.701	0.722	0.690	0.671	0.659	0.692	0.692	0.678	0.668	0.722
5	0.600	0.578	0.534	0.541	0.536	0.518	0.499	0.512	0.550	0.600
6	0.593	0.585	0.569	0.588	0.595	0.630	0.622	0.641	0.657	0.657
7	0.529	0.572	0.604	0.562	0.582	0.564	0.510	0.483	0.469	0.604
col BH										
	0.804	0.810	0.828	0.848	0.834	0.823	0.824	0.814	0.800	0.848

Notes: See notes to Table 1.

Table 15
Stock Returns and Macroeconomic Announcements
Retrospective weak G non-causality test: CI Test Regression 3
(ψ, ψ_y, ψ_d: ridgelet function)

$\tau \diagdown q$	1	2	3	4	5	6	7	8	9	row BH
0	0.006	0.006	0.009	0.011	0.011	0.013	0.011	0.011	0.011	0.013
1	0.002	0.002	0.002	0.002	0.003	0.003	0.004	0.003	0.002	0.004
2	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.002	0.001	0.002
3	0.003	0.004	0.007	0.003	0.004	0.006	0.006	0.005	0.005	0.007
4	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.001	0.001	0.002
5	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.002
6	0.001	0.002	0.002	0.001	0.002	0.001	0.001	0.001	0.001	0.002
7	0.003	0.003	0.003	0.003	0.002	0.003	0.004	0.003	0.003	0.004
col BH										
	0.006	0.006	0.008	0.006	0.008	0.007	0.007	0.006	0.005	0.013

Notes: See notes to Table 1.

8 Appendix: Proofs

Proof of Proposition 3.2: (a) By Assumption A.1(a), $D_t = c_{2,t}(D_{0,}W^{t-1}, V^{t-1}, Z^{t-1})$, so D^t is some function of D_0,$W^{t-1}, V^{t-1}, Z^{t-1}$. For any constant c, we have $Y_{t+k} \perp c \mid D_0, W^{t-1}, V^{t-1}, Z^{t-1}$. Then by D lemma 4.1 $(Y_{t+k}, D_0, W^{t-1}, V^{t-1}, Z^{t-1}) \perp (c, D_0, W^{t-1}, V^{t-1}, Z^{t-1}) \mid D_0, W^{t-1}, V^{t-1}, Z^{t-1}$. Two applications of D lemma 4.2 then give $Y_{t+k} \perp D^t \mid D_0, W^{t-1}, V^{t-1}, Z^{t-1}$. Similarly, as D^t is also a function of $D_0, Y^t, W^{t-1}, V^{t-1}, Z^{t-1}$, we have $Y_{t+k} \perp D^t \mid D_0, Y^t, W^{t-1}, V^{t-1}, Z^{t-1}$. (b) As $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$, there exists a measurable function f_{t+k} such that $Y_{t+k} = f_{t+k}(Y_0, V^{t+k-1}, Z^{t+k-1})$. Substituting $V_{t+k-1} = c_{0,t+k-1}(V_0, Z^{t+k-1})$ yields a measurable function g_{t+k} such that $Y_{t+k} = g_{t+k}(Y_{0,}V_{0,}Z^{t+k-1})$. By the same argument as in (a), we have $Y_{t+k} \perp D^t \mid Y_{0,}V_{0,}Z^{t+k-1}$. As $\sigma(Y_{0,}V_{0,}Z^{t+k-1})$, $\sigma(Y_0, V^t, Z^{t+k-1})$, and $\sigma(Y^t, V^t, Z^{t+k-1})$ contain the same information, it follows that $Y_{t+k} \perp D^t \mid Y_0, V^t, Z^{t+k-1}$ and $Y_{t+k} \perp D^t \mid Y^t, V^t, Z^{t+k-1}$. ■

Proof of Proposition 3.4: (a) By D lemma 4.1, $D^t \perp U^t \mid Y_0, X^t$ (Assumption A.2(a)) implies $(D^t, Y_0, X^t) \perp (U^t, Y_0, X^t) \mid Y_0, X^t$. Given Assumption A.1(a) and $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$, there exists a measurable function f_{t+1} such that $Y_{t+1} = f_{t+1}(Y_0, \tilde{V}^t, \tilde{Z}^t, U^t)$. Because $X^t = (\tilde{V}^t, W^t, \tilde{Z}^t)$, it follows from lemma 4.2(i) of D that $f_{t+1}(Y_0, \tilde{V}^t, \tilde{Z}^t, U^t) \perp D^t \mid Y_0, X^t$. That is, $Y_{t+1} \perp D^t \mid Y_0, X^t$. Next, $D^t \not\Rightarrow_{\mathcal{S}} Y_{t+1}$ and $(D^t, Y_0, X^t) \perp (U^t, Y_0, X^t) \mid Y_0, X^t$ imply $(D^t, Y_0, X^t) \perp (U^t, Y_0, X^t) \mid Y^t, X^t$ by D lemma 4.2(ii), as $Y_1^t \equiv (Y_1, ..., Y_t) = g_1^t(U^t, Y_0, X^t)$, say, for some function g_1^t. It follows from lemma 4.2(i) of D that $Y_{t+1} \perp D^t \mid Y^t, X^t$. ■

Proof of Proposition 3.6: The argument is entirely parallel to Proposition 3.4, replacing X^t with X^T. ■

Proof of Theorem 3.8: For part (a), we show that for the sets specified in A.3 (a), $P[Y_{t+1} \in B_Y | D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \neq P[Y_{t+1} \in B_Y | Y_0 \in B_0, X^t \in B_X]$, as this implies $Y_{t+1} \not\perp D^t \mid Y_0, X^t$. Because A.3(a.i) ensures $P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X] > 0$ and $P[Y_0 \in B_0, X^t \in B_X] > 0$, we have

$$P[\ Y_{t+1} \ \in \ B_Y \ | D^t \in B_D, Y_0 \in B_0, X^t \in B_X]$$

$$= \frac{P[Y_{t+1} \in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X]}{P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X] P[Y_0 \in B_0, X^t \in B_X]}$$

and

$$\begin{aligned} P[Y_{t+1} &\in B_Y | Y_0 \in B_0, X^t \in B_X] \\ &= \frac{P[Y_{t+1} \in B_Y, Y_0 \in B_0, X^t \in B_X]}{P[Y_0 \in B_0, X^t \in B_X]}. \end{aligned}$$

By the law of total probability

$$\begin{aligned} &P[Y_{t+1} \in B_Y, Y_0 \in B_0, X^t \in B_X] \\ &\qquad = P[Y_{t+1} \in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \\ &\qquad\qquad + P[Y_{t+1} \in B_Y, D^t \notin B_D, Y_0 \in B_0, X^t \in B_X]. \end{aligned}$$

But A.3(a.iii) implies that

$$P[Y_{t+1} \in B_Y, D^t \notin B_D, Y_0 \in B_0, X^t \in B_X] = 0,$$

so

$$P[Y_{t+1} \in B_Y, Y_0 \in B_0, X^t \in B_X] = P[Y_{t+1} \in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X].$$

Because $P[Y_0 \in B_0, X^t \in B_X] > 0$ and $P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X] > 0$, we have $P[Y_{t+1} \in B_Y | D^t \in B_D, Y_0 \in B_0, X^t \in B_X] = P[Y_{t+1} \in B_Y | Y_0 \in B_0, X^t \in B_X]$ if and only if

$$\begin{aligned} P[Y_{t+1} &\in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \\ &= P[Y_{t+1} \in B_Y, Y_0 \in B_0, X^t \in B_X] \; P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X]. \end{aligned}$$

It follows that $P[Y_{t+1} \in B_Y | D^t \in B_D, Y_0 \in B_0, X^t \in B_X] = P[Y_{t+1} \in B_Y | Y_0 \in B_0, X^t \in B_X]$ if and only if

$$\begin{aligned} P[Y_{t+1} &\in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \\ &= P[Y_{t+1} \in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \; P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X]. \end{aligned}$$

This holds if and only if either

$$P[Y_{t+1} \in B_Y, D^t \in B_D, Y_0 \in B_0, X^t \in B_X] = 0$$

or

$$P[D^t \in B_D | Y_0 \in B_0, X^t \in B_X] = 1.$$

But these possibilities are ruled out by A.3(a.i) and A.3(a.ii) respectively. It follows that

$$\begin{aligned} P[Y_{t+1} &\in B_Y | D^t \in B_D, Y_0 \in B_0, X^t \in B_X] \\ &\neq P[Y_{t+1} \in B_Y | Y_0 \in B_0, X^t \in B_X], \end{aligned}$$

and the proof of part (a) is complete. The proof of part (b) is similar. ■

Proof of Proposition 4.1: For part (a), using the same argument as in Proposition 4.3 of White and Kennedy (2008), we obtain

$$\begin{aligned} dF_{1,0}(y_1|y_0, d^0, x^0) &= dF_{-\tau}(y_1|y_0, d_0, x^0_{-\tau}) && (t=0) \\ dF_{t+1,t}(y_{t+1}|y_0, d^t, x^t) &= \int dF_{-\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau})\ dF_{t,t-1}(y_t|y_0, d^{t-1}, x^{t-1}) \\ & && t = 1, \dots . \end{aligned}$$

Then by applying A.5, we have

$$\begin{aligned} dF_{1,0}(y_1|y_0, d^0, x^0) &= dF_{-\tau}(y_1|y_0, d_0, x^0_{-\tau}) && (t=0) \\ dF_{t+1,t}(y_{t+1}|y_0, d^t, x^t) &= \int dF_{-\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau})\ dF_{t,t-1}(y_t|y_0, d^{t-1}, x^{t-1}) \\ & && t = 1, \dots . \end{aligned}$$

The proof of part (b) exactly follows Proposition 4.3 of White and Kennedy (2008). ■

Proof of Proposition 4.2: For part(a), first we show necessity. Suppose $Y_{t+1} \perp D_t \mid Y_t, X^t_{t-\tau}$ for all $t = 0, 1...$. Take $t = 0$. Then by Proposition 4.1(a), we have

$$\begin{aligned} dF_{1,0}(y_1|y_0, d^0, x^0) &= dF_{-\tau}(y_1|y_0, d_0, x^0_{-\tau}) \\ &= dF_{-\tau}(y_1|y_0, x^0_{-\tau}), \end{aligned}$$

where the second equation follows from the maintained hypothesis. We also have for $t = 1, ...$

$$dF_{\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau}) = dF_{-\tau}(y_{t+1}|y_t, x^t_{t-\tau}).$$

By the maintained hypothesis, the recursions of Proposition 4.1(a) give

$$\begin{aligned} dF_{t+1,t}(y_{t+1}|y_0, d^t, x^t) &= \int dF_{-\tau}(y_{t+1}|y_t, d_t, x^t_{t-\tau}) dF_{t,t-1}(y_t|y_0, d^{t-1}, x^{t-1}) \\ &= \int dF_{-\tau}(y_{t+1}|y_t, x^t_{t-\tau}) dF_{t,t-1}(y_t|y_0, x^{t-1}), \\ &= dF_{t+1,t}(y_{t+1}|y_0, x^t) \end{aligned}$$

$$t = 1, \dots .$$

so that necessity is proved.

For sufficiency, we give a proof by contradiction. Suppose there exists some t_0 such that $Y_{t_0+1} \not\perp D_{t_0} | Y_{t_0}, X_{t_0-\tau}^{t_0}$. This ensures that $dF_{-\tau}(y_{t_0+1}|y_{t_0}, d_{t_0}, x_{t_0-\tau}^{t_0})$ must be a non-constant function of d_{t_0} on a set of positive probability. It follows that $dF_{-\tau}(y_1|y_0, d_0, x_{-\tau}^0)$ must be a non-constant function of d_0 on a set of positive probability, since we assume the conditional density $dF_{-\tau}$ is the same for all t. By Proposition 4.1(a) for $t = 0$, we have

$$dF_{1,0}(y_1|y_0, d^0, x^0) = dF_{-\tau}(y_1|y_0, d_0, x_{-\tau}^0).$$

This implies that $dF_{1,0}(y_1|y_0, d^0, x^0)$ is a non-constant function of d^0 on a set of positive probability. But this contradicts $Y_{t+1} \perp D^t \mid Y_0, X^t$ when $t = 0$. The sufficiency part is proved and the proof of part (a) is complete. The proof of part (b) is similar. ■

Proof of Proposition 5.1: For part (a), A.6(a.ii), i.e., $D^t \perp (\ddot{U}^t, \tilde{W}_0) \mid Y_0, U^t, X^t$, and $D^t \perp U^t \mid Y_0, X^t$ imply

$$(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid Y_0, X^t \tag{A.i}$$

by D Lemma 4.3. $(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid Y_0, X^t$ implies $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid Y_0, X^t$ by D Lemma 4.1. Since $\tilde{W}_{t+1} = b_{3,t+1}(\tilde{W}^t, X^t, U^t, \ddot{U}^t)$, by recursive substitution, there exists some measurable function f_{t+1} such that $\tilde{W}_{t+1} = f_{t+1}(\tilde{W}_0, X^t, U^t, \ddot{U}^t)$. Hence $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid Y_0, X^t$ implies $(\tilde{W}_{t+1}, \tilde{W}_0) \perp D^t \mid Y_0, X^t$ by D Lemma 4.2(i). $(\tilde{W}_{t+1}, \tilde{W}_0) \perp D^t \mid Y_0, X^t$ implies $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, Y_0, X^t$ by D Lemma 4.2(ii). Further, A.7(a) implies $\tilde{W}_{t+1} \perp Y_0 \mid \tilde{W}_0, X^t$ by D Lemma 4.2. $\tilde{W}_{t+1} \perp Y_0 \mid \tilde{W}_0, X^t$ and $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, Y_0, X^t$ together imply $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, X^t$ by D Lemma 4.3. The proof of part(a) is complete. The proof of part(b) is parallel, replacing X^t with X^T. ■

Proof of Proposition 5.2: For part (a), we first show that $dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_0, d^t, x^t)$ has the following recursive representation:

$$dF_{1,0}(\tilde{w}_1|\tilde{w}_0, d^0, x^0) = dF_{-\tau}(\tilde{w}_1|\tilde{w}_0, d_0, x_{-\tau}^0) \qquad t = 0 \tag{A.ii}$$

$$dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_0, d^t, x^t) = \int dF_{-\tau}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x_{t-\tau}^t)\, dF_{t,t-1}(\tilde{w}_t|\tilde{w}_0, d^{t-1}, x^{t-1}) \qquad t = 1, \dots . \tag{A.iii}$$

Eq.(A.ii) follows directly from A.8(a) when $t = 0$. Viewing $\tilde{W}_t$ as a state vector, we have the prediction density equation

$$dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_0, d^t, x^t) = \int dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_t, \tilde{w}_0, d^t, x^t)\, dF_{t,t-1}(\tilde{w}_t|\tilde{w}_0, d^t, x^t).$$

To verify eq.(A.iii), it suffices to verify eqs.(A.iv), (A.v), and (A.vi):

$$dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_t, \tilde{w}_0, d^t, x^t) = dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x^t) \quad \text{(A.iv)}$$

$$= dF_{-\tau}(\tilde{w}_{t+1}|\tilde{w}_t, d_t, x^t_{t-\tau}) \quad \text{(A.v)}$$

$$dF_{t,t-1}(\tilde{w}_t|\tilde{w}_0, d^t, x^t) = dF_{t,t-1}(\tilde{w}_t|\tilde{w}_0, d^{t-1}, x^{t-1}). \quad \text{(A.vi)}$$

Eq.(A.v) is A.6(a). Eq.(A.iv) says that $\tilde{W}_{t+1} \perp (\tilde{W}_{0,} D^{t-1}) \mid \tilde{W}_t, D_t, X^t$. We now verify this. Now $(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid Y_0, X^t$ (eq.(A.i)) implies $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ by D Lemma 4.1. Since $(\tilde{W}_{t+1}, \tilde{W}_t)$ is a function of $(\tilde{W}_0, U^t, \ddot{U}^t, X^t)$ by recursive substitution, $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ implies $(\tilde{W}_{t+1}, \tilde{W}_t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ by D Lemma 4.2(i). $(\tilde{W}_{t+1}, \tilde{W}_t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ implies $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, \tilde{W}_t, Y_0, X^t$ by D Lemma 4.2(ii). A.7(a) implies $\tilde{W}_{t+1} \perp (Y_0, \tilde{W}_0) \mid \tilde{W}_t, X^t$ by D Lemma 4.2(ii). $\tilde{W}_{t+1} \perp (Y_0, \tilde{W}_0) \mid \tilde{W}_t, X^t$ implies $\tilde{W}_{t+1} \perp Y_0 \mid \tilde{W}_0, \tilde{W}_t, X^t$ by D Lemma 4.2(ii). $\tilde{W}_{t+1} \perp Y_0 \mid \tilde{W}_0, \tilde{W}_t, X^t$ and $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, \tilde{W}_t, Y_0, X^t$ imply $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, \tilde{W}_t, X^t$ by D Lemma 4.3. A.7(a) also implies $\tilde{W}_{t+1} \perp \tilde{W}_0 \mid \tilde{W}_t, X^t$ by D Lemma 4.2. $\tilde{W}_{t+1} \perp \tilde{W}_0 \mid \tilde{W}_t, X^t$ and $\tilde{W}_{t+1} \perp D^t \mid \tilde{W}_0, \tilde{W}_t, X^t$ imply $\tilde{W}_{t+1} \perp (\tilde{W}_0, D^t) \mid \tilde{W}_t, X^t$ by D Lemma 4.3. $\tilde{W}_{t+1} \perp (\tilde{W}_0, D^t) \mid \tilde{W}_t, X^t$ implies $\tilde{W}_{t+1} \perp (\tilde{W}_0, D^t) \mid \tilde{W}_t, D_t, X^t$ by D Lemma 4.2(ii). Further $\tilde{W}_{t+1} \perp (\tilde{W}_0, D^t) \mid \tilde{W}_t, D_t, X^t$ implies $\tilde{W}_{t+1} \perp (\tilde{W}_0, D^{t-1}) \mid \tilde{W}_t, D_t, X^t$ by D Lemma 4.2(i). This verifies eq.(A.iv).

Eq.(A.vi) says that $\tilde{W}_t \perp (D_t, X_t) \mid \tilde{W}_0, D^{t-1}, X^{t-1}$. We now verify this. Now $(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid Y_0, X^t$ (eq.(A.i)) implies $(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ by D Lemma 4.2(ii). $(U^t, \ddot{U}^t, \tilde{W}_0) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ implies $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ by D Lemma 4.1. $(U^t, \ddot{U}^t, \tilde{W}_0, X^t) \perp D^t \mid \tilde{W}_0, Y_0, X^t$ implies $\tilde{W}_t \perp D^t \mid \tilde{W}_0, Y_0, X^t$, since $\tilde{W}_t$ is a function of $(U^t, \ddot{U}^t, \tilde{W}_0, X^t)$ by recursive substitution. A.7(a) implies $\tilde{W}_t \perp Y_0 \mid \tilde{W}_0, X^t$ by D Lemma 4.2. $\tilde{W}_t \perp Y_0 \mid \tilde{W}_0, X^t$ and $\tilde{W}_t \perp D^t \mid \tilde{W}_0, Y_0, X^t$ imply $\tilde{W}_t \perp D^t \mid \tilde{W}_0, X^t$ by D Lemma 4.3. $\tilde{W}_t \perp D^t \mid \tilde{W}_0, X^t$ and A.9 imply $\tilde{W}_t \perp (D^t, X_t) \mid \tilde{W}_0, X^{t-1}$ by D Lemma 4.3. $\tilde{W}_t \perp (D^t, X_t) \mid \tilde{W}_0, X^{t-1}$ implies $\tilde{W}_t \perp (D_t, X_t) \mid \tilde{W}_0, D^{t-1}, X^{t-1}$ by D Lemma 4.2. This verifies eq.(A.vi).

With the recursive representation of $dF_{t+1,t}(\tilde{w}_{t+1}|\tilde{w}_0, d^t, x^t)$, the rest of the proof of part (a) follows the proof of Proposition 4.2(a) exactly. The proof of part (a) is thus complete. The proof of part (b) is analogous. ■

Proof of Proposition 5.3: We first define the null hypothesis H_0 and alternative hypothesis H_A. We denote (retrospective) conditional exogeneity H_1 and (retrospective) (weak) G non-causality H_2. Then H_0 and H_A are defined as

$$\begin{aligned} H_0 &: \text{ Either } H_1 \text{ does not hold or } H_2 \text{ holds.} \\ H_A &: \text{ } H_1 \text{ holds and } H_2 \text{ does not hold.} \end{aligned}$$

Under H_0, we fail to reject structural non-causality. Under H_A, we reject structural non-causality. Let P_0 denote the true probability measure under H_0. There are three cases under H_0.

Case 1: H_1 holds and H_2 holds. In this case,

$$\begin{aligned} \alpha &= P_0\left[\text{fail to reject } H_1 \text{ and reject } H_2\right] \\ &\leq \min\left\{P_0\left[\text{fail to reject } H_1\ \right], P_0\left[\text{reject } H_2\right]\right\} \\ &= \min\{1-\alpha_1, \alpha_2\}; \text{ and} \\ \alpha &= 1 - P_0\left[\text{reject } H_1 \text{ or fail to reject } H_2\right] \\ &\geq 1 - P_0\left[\text{reject } H_1\right] - P_0\left[\text{fail to reject } H_2\right] \\ &= 1 - \alpha_1 - (1-\alpha_2) = \alpha_2 - \alpha_1. \end{aligned}$$

Case 2: H_1 does not hold and H_2 does not hold. Similarly, in this case,

$$\begin{aligned} \alpha &\leq \min\left\{P_0\left[\text{fail to reject } H_1\right], P_0\left[\text{reject } H_2\right]\right\} \\ &= \min\{1-\pi_1, \pi_2\}; \text{ and} \\ \alpha &\geq 1 - P_0\left[\text{reject } H_1\right] - P_0\left[\text{fail to reject } H_2\right] \\ &= 1 - \pi_1 - (1-\pi_2) = \pi_2 - \pi_1. \end{aligned}$$

Case 3: H_1 does not hold and H_2 holds. Similarly, in this case,

$$\begin{aligned} \alpha &\leq \min\left\{P_0\left[\text{fail to reject } H_1\ \right], P_0\left[\text{reject } H_2\right]\right\} \\ &= \min\{1-\pi_1, \alpha_2\}; \text{ and} \\ \alpha &\geq 1 - P_0\left[\text{reject } H_1\right] - P_0\left[\text{fail to reject } H_2\right] \\ &= 1 - \pi_1 - (1-\alpha_2) = \alpha_2 - \pi_1. \end{aligned}$$

Combining the three cases gives

$$\begin{aligned} \alpha &\leq \max\{\min\{1-\alpha_1, \alpha_2\}, \min\{1-\pi_1, \pi_2\}, \min\{1-\pi_1, \alpha_2\}\} \text{ and} \\ \alpha &\geq \max\left\{0, \min\left\{\left(\alpha_2-\alpha_1\right), \left(\pi_2-\pi_1\right), \left(\alpha_2-\pi_1\right)\right\}\right\}. \end{aligned}$$

Let P_A denote the probability measure under $H_{A.}$ Then

$$\begin{aligned}\pi &= 1-P_A\,[\text{Either reject } H_1 \text{ or fail to reject } H_2]\\ &\geq 1-P_A\,[\text{reject } H_1]-\ P_A\,[\text{fail to reject } H_2]\\ &= 1-\alpha_1-(1-\pi_2)=\pi_2-\alpha_1;\ \text{and}\\ \pi &= P_A\,[\text{fail to reject } H_1 \text{ and reject } H_2]\\ &\leq \min\{P_A\,[\text{fail to reject } H_1]\,,\ P_A\,[\text{reject } H_2]\}\\ &= \min\{1-\alpha_1,\pi_2\}\,.\end{aligned}$$

Hence,

$$\pi_2-\alpha_1\leq\pi\leq\min\{1-\alpha_1,\pi_2\}\,.\ \blacksquare$$

Proof of Proposition 5.4: From Proposition 5.3, we have $\alpha_T\leq\max\{\min\{1-\alpha_{1T},\alpha_{2T}\},\min\{1-\pi_{1T},\pi_{2T}\},\min\{1-\pi_{1T},\alpha_{2T}\}\}$, which implies

$$\begin{aligned}\limsup\alpha_T &\leq \limsup\{\max\{\min\{1-\alpha_{1T},\alpha_{2T}\},\min\{1-\pi_{1T},\pi_{2T}\},\min\{1-\pi_{1T},\alpha_{2T}\}\}\}\\ &= \lim\{\max\{\min\{1-\alpha_{1T},\alpha_{2T}\},\min\{1-\pi_{1T},\pi_{2T}\},\min\{1-\pi_{1T},\alpha_{2T}\}\}\}\\ &= \max\{\min\{1-\alpha_1,\alpha_2\}\,,\min\{0,1\}\,,\min\{0,\alpha_2\}\}=\min\{1-\alpha_1,\alpha_2\}\,.\end{aligned}$$

Also from Proposition 5.3, we have $\alpha_T\geq\{\max 0,\min\{(\alpha_{2T}-\alpha_{1T})\,,(\pi_{2T}-\pi_{1T})\,,(\alpha_{2T}-\pi_{1T})\}\}\,,$ which implies

$$\begin{aligned}\liminf\alpha_T &\geq \liminf\{\max\{0,\min\{(\alpha_{2T}-\alpha_{1T})\,,(\pi_{2T}-\pi_{1T})\,,(\alpha_{2T}-\pi_{1T})\}\}\}\\ &= \lim\{\max\{0,\min\{(\alpha_{2T}-\alpha_{1T})\,,(\pi_{2T}-\pi_{1T})\,,(\alpha_{2T}-\pi_{1T})\}\}\}\\ &= \max\{0,\min\{(\alpha_2-\alpha_1)\,,0,(\alpha_2-1)\}\}=0.\end{aligned}$$

Thus,

$$0\leq\liminf\alpha_T\leq\limsup\alpha_T\leq\min\{1-\alpha_1,\alpha_2\}\,.$$

From Proposition 5.3, we have $\pi_T\leq\min\{1-\alpha_{1T},\pi_{2T}\}$, which implies

$$\begin{aligned}\limsup\pi_T &\leq \limsup\{\min\{1-\alpha_{1T},\pi_{2T}\}\}\\ &= \lim\{\min\{1-\alpha_{1T},\pi_{2T}\}\}\\ &= \min\{1-\alpha_1,1\}=1-\alpha_1.\end{aligned}$$

Also from Proposition 5.3, we have $\pi_T\geq\pi_{2T}-\alpha_{1T}$, which implies

$$\begin{aligned}\liminf\pi_T &\geq \liminf(\pi_{2T}-\alpha_{1T})\\ &= \lim(\pi_{2T}-\alpha_{1T})\\ &= 1-\alpha_1.\end{aligned}$$

Thus,

$$1 - \alpha_1 \leq \liminf \pi_T \leq \limsup \pi_T \leq 1 - \alpha_1.$$

This further implies

$$\lim \pi_T = 1 - \alpha_1. \quad \blacksquare$$

References

[1] Angrist, J. and G. Kuersteiner (2004), "Semiparametric Causality Tests Using the Policy Propensity Score," NBER Working Paper No. 10975.

[2] Borenstein, S., A. Cameron, and R. Gilbert (1997), "Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes," *Quarterly Journal of Economics*, 112, 305-339.

[3] Candès, E. (1999), "Ridgelets: Estimating with Ridge Functions," *Annals of Statistics*, 31, 1561-1599.

[4] Chalak, K. and H. White (2007a), "Independence and Conditional Independence in Causal Systems," UCSD Dept. of Economics Discussion Paper.

[5] Chalak, K. and H. White (2007b), "Identification with Conditioning Instruments in Causal Systems," UCSD Dept. of Economics Discussion Paper.

[6] Chalak, K. and H. White (2007c), "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," UCSD Dept. of Economics Discussion Paper.

[7] Chen, N., R. Roll and S. Ross (1986), "Economic Forces and the Stock Market," *Journal of Business*, 59, 383-403.

[8] Chow, G. (1997). *Dynamic Economics: Optimization by the Lagrange Method*. Oxford: Oxford University Press.

[9] Dawid, A.P. (1979), "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society*, Series B, 41, 1-31.

[10] Delgado, M. A. and W. Gonzalez-Manteiga (2001), "Significance Testing in Nonparametric Regression Based on the Bootstrap," *Annals of Statistics*, 29, 1469-1507.

[11] Fernandes, M. and R. G. Flores (2001), "Tests for Conditional Independence, Markovian Dynamics, and Noncausality," Discussion Paper, European University Institute.

[12] Fisher, F. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill.

[13] Fisher, F. (1970), "A Correspondence Principle for Simultaneous Equations Models," *Econometrica*, 38, 73-92.

[14] Flannery, M. and A. Protopapadakis, 2002, "Macroeconomic Factors Do Influence Aggregate Stock Returns," *Review of Financial Studies*, 15, 751–782.

[15] Florens, J.P. and D. Fougère (1996), "Non-causality in Continuous Time," *Econometrica*, 64, 1195–1212.

[16] Florens, J.P. and M. Mouchart (1982), "A Note on Non-causality," *Econometrica*, 50, 583–591.

[17] Goldberger, A. (1972), "Structural Equation Methods in the Social Sciences," *Econometrica*, 40, 979-1001.

[18] Granger, C.W.J. (1969), "Investigating Causal Relations by Econometric Models and Cross-spectral Methods," *Econometrica*, 37, 424–438.

[19] Granger, C.W.J. (1980), "Testing for Causality, a Personal Viewpoint," *Journal of Economic Dynamics and Control*, 2, 329-352.

[20] Granger, C.W.J. (1988), "Some Recent Developments in a Concept of Causality," *Journal of Econometrics*, 39, 199-211.

[21] Granger, C.W.J. and P. Newbold (1986). *Forecasting Economic Time Series*. New York: Academic Press (2nd edition).

[22] Heckman, J. (2008), "Econometric Causality," NBER Working Paper No. 13934.

[23] Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 75, 800–802.

[24] Leeper, E. (1997), "Narrative and VAR approaches to Monetary Policy: Common Identification Problems," *Journal of Monetary Economics*, 40, 641-657.

[25] Linton, O. and P. Gozalo (1997), "Conditional Independence Restrictions: Testing and Estimation," Discussion Paper, Cowles Foundation for Research in Economics, Yale University.

[26] Newey, W. K. and K. D. West (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

[27] Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

[28] Romer, C. and D. Romer (1989), "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, 4, 121-170.

[29] Romer, C. and D. Romer (1994), "Monetary Policy Matters," *Journal of Monetary Economics*, 34, 75-88.

[30] Shapiro, M. (1994), "Federal Reserve Policy: Cause and Effect," in G. Mankiw, ed., *Monetary Policy*. Chicago: University of Chicago Press, 307-334.

[31] Sims, C. (1972), "Money, Income, and Causality," *American Economic Review*, 62, 540-552.

[32] Strotz, R. and H. Wold (1960), "Recursive vs. Nonrecursive Systems: An Attempt at Synthesis," *Econometrica*, 28, 417-427.

[33] Su, L. and H. White (2007a), "A Consistent Characteristic Function-Based Test for Conditional Independence," *Journal of Econometrics*, 141, 807-834.

[34] Su, L. and H. White (2007b), "Testing Conditional Independence via Empirical Likelihood," UCSD Dept. of Economics Discussion Paper.

[35] Su, L. and H. White (2008), "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory*, 24, 829–864.

[36] White, H. (2006a), "Time-series Estimation of the Effects of Natural Experiments," *Journal of Econometrics*, 135, 527-566.

[37] White, H (2006b), "Approximate Nonlinear Forecasting Methods," in G. Elliott, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*. New York: Elsevier, 460-512.

[38] White, H. and K. Chalak (2007a), "Identifying Effects of Endogenous Causes in Nonseparable Systems Using Covariates," UCSD Dept. of Economics Discussion Paper.

[39] White, H. and K. Chalak (2007b), "Settable Systems: An Extension of Pearl's Causal Model with Optimization, Equilibrium, and Learning," UCSD Dept. of Economics Discussion Paper.

[40] White, H. and K. Chalak (2007c), "Parametric and Nonparametric Estimation of Covariate-Conditioned Average Effects," UCSD Dept. of Economics Discussion Paper.

[41] White, H. and P. Kennedy (2008), "Retrospective Estimation of Causal Effects Through Time," in J. Castle and N. Shephard, eds., *A Festschrift in Honour of David Hendry*. Oxford: Oxford University Press (forthcoming).