A nonparametric dynamic causal model for macroeconometrics*

Ashesh Rambachan
Department of Economics,
Harvard University
asheshr@g.harvard.edu

Neil Shephard
Department of Economics and
Department of Statistics,
Harvard University
shephard@fas.harvard.edu

March 1, 2019

Abstract

This paper uses potential outcome time series to provide a nonparametric framework for quantifying dynamic causal effects in macroeconometrics. This provides sufficient conditions for the nonparametric identification of dynamic causal effects as well as clarify the causal content of several common assumptions and methods in macroeconomics. Our key identifying assumption is shown to be *non-anticipating treatments* which enables nonparametric inference on dynamic causal effects. Next, we provide a formal definition of a "shock" and this leads to a *shocked potential outcome time series*. This is a nonparametric statement of the Frisch-Slutzky paradigm. The common additional assumptions that the causal effects are additive and that the treatments are shocks place substantial restrictions on the underlying dynamic causal estimands. We use this structure to causally interpret several common estimation strategies. We provide sufficient conditions under which local projections is causally interpretable and show that the standard assumptions for local projections with an instrument are not sufficient to identify dynamic causal effects. We finally show that the structural vector moving average form is causally equivalent to a restricted potential outcome time series under the usual invertibility assumption.
Keywords: Causality, nonparametric, potential outcomes, shocks, time series.

*We thank Iavor Bojinov, Gary Chamberlain, Fabrizia Mealli, James M. Robins and James H. Stock for conversations that have developed our thinking on causality. We are also grateful to Isaiah Andrews, John Campbell, Peng Ding, Avi Feller, Ron Gallant, Peter Hansen, Kosuke Imai, Guido Kuersteiner, Daniel Lewis, Luke Miratrix, Sendhil Mullainathan, Susan A. Murphy, Andrew Patton, Natesh Pillai and Mikkel Plagborg-Moller for valuable comments and feedback. All remaining errors are our own.

1 Introduction

Many of the most important empirical questions in macroeconomics center on dynamic causal effects. A dynamic causal effect traces out the effect of some policy intervention or change in economic conditions on observed macroeconomic variables over time. For example, suppose the Federal Open Market Committee raises the target Federal funds rate by 25 basis points. How will this affect unemployment over the next several quarters? To answer this type of question, empirical macroeconomists typically rely on models that are estimated on aggregate time series data. In this paper, we provide a nonparametric framework for causal inference on macroeconomic time series data.

Conceptualizing and estimating dynamic causal effects from time series data is challenging. Dynamic feedback between the treatments and observed outcomes makes it difficult to disentangle causes from effects. Therefore explicit assumptions about the timing of such responses are often necessary (Sims, 1972, 1980). Additionally, the econometrician usually only observes several decades of monthly time series data and so, obtaining precise estimates is difficult. Given these challenges, much of the literature on dynamic causal effects in macroeconomics typically relies on parametric linear models. Canonical examples are structural vector autoregressions (Sims, 1980) and local projections (Jordá, 2005).[1] While tractable, the heavy emphasis on linear models has drawbacks. The strong functional form assumptions may implicitly impose substantial restrictions on the underlying dynamic causal effects. If so, are these restrictions plausible in the empirical applications of interest? Moreover, the role of particular sets of assumptions are often unclear in existing approaches. Is an assumption needed to define a dynamic causal effect, identify a dynamic causal effect or simplify estimation? For example, it is common in macroeconomics to restrict attention to the causal effects of "shocks" (Frisch, 1933; Slutzky, 1937; Ramey, 2016). Is this a convenient choice or does it reflect something deeper?

We address each of these questions by building upon the nonparametric *potential outcome time series* of Bojinov and Shephard (2019), which sought to analyze randomized experiments carried out on time series. We extend this to common observational settings. Our setting is familiar to macroeconomists. We have a time series observed over $t = 1, \ldots, T$ periods made up of a vector of treatments W_t and a vector of outcomes Y_t. Following the cross-sectional tradition (Imbens and Wooldridge (2009); Imbens and Rubin (2015)), there is a vector of potential outcomes associated with each outcome that describe what would have been observed along alternative dynamic treatment

[1] Structural vector autoregressions are typically motivated as a linear approximation to an equilibrium arising from an underlying dynamic stochastic general equilibrium model such as Christiano et al. (1999, 2005), Smets and Wouters (2003, 2007). However, there are exceptions such as Priestley (1988), Engle et al. (1990), Gallant et al. (1993), Koop et al. (1996) that directly examine the impulse response functions of non-linear, multivariate time series models.

paths. A *dynamic causal effect* is then generically defined as the comparison of potential outcomes along different treatment paths at a fixed point in time. We believe that the potential outcome time series framework provides a flexible, fully nonparametric foundation upon which to build new methods and interpret existing methods for causal inference on time series.

With this as our starting point, we make four novel contributions to the macroeconometric literature. First, we introduce a variety of new concepts that deepen our understanding of dynamic causal effects and suggest new ways of thinking about them. Many of these will extend ideas that are well understood in cross-sectional work to the time series setting. For example, we discuss additional restrictions on dynamic causal effects such as additivity that play an enormous role in existing empirical work in macroeconomics. We also introduce several dynamic causal estimands. The estimands differ based upon what is treated as random and therefore averaged over. For example, what we call the "weighted causal effect" averages differences of potential outcomes over the random assignment of the treatment paths holding the potential outcomes fixed, whereas the "causal response function" averages over both the random treatment paths as well as the potential outcomes.

This distinction is a novel conceptual contribution. In particular, researchers currently attempt to make causal statements that hold in all periods – e.g. "a 25 basis point increase in the Federal funds rate causes unemployment to rise by 0.5 percentage points." This is captured in the literature's primary focus on the impulse response function. Using microeconometric nomenclature, these are statements about *superpopulation* causal effects. We introduce a tractable notion of a *sample* causal effect into time series that involves no out-of-sample extrapolation – e.g. "a 25 basis point increase in the Federal funds rate caused unemployment to rise on average by 0.5 percentage points in the period over which we have data." This subtle difference is commonly discussed in microeconometrics (Aronow and Samii, 2016; Abadie et al., 2018). To the best of our knowledge, we are the first to introduce this distinction into time series macroeconometrics.

Second, we provide sufficient conditions for the nonparametric identification of these dynamic causal estimands. The identifying assumption is that the treatments are *non-anticipating*. That is, the assignment mechanism of the current treatment does not depend on future potential outcomes conditional on past outcomes and treatments. When this does not hold, inference on dynamic causal effects becomes difficult as treatment in the current period is correlated with future outcomes and so, comparisons of observed outcomes along different treatment paths is confounded. In this sense, we think of non-anticipation as the time series version of unconfoundedness. When the treatments are non-anticipating, we show that we can construct nonparametric estimators that are unbiased,

consistent, asymptotically normally distributed and can be used to provide exact testing of sharp nulls or asymptotically conservative tests of more general hypotheses.

Third, we make substantial progress towards clarifying the causal content of common assumptions in applied macroeconometrics. As mentioned, macroeconomists typically focus on the dynamic causal effects of shocks (Frisch, 1933; Slutzky, 1937). To understand why, we provide a definition of a shock in the potential outcome time series framework that formalizes the popular heuristic that shocks are "unpredictable" (Ramey, 2016; Stock and Watson, 2016). We thereby provide a causal nonparametric formalization of the dominant Frisch (1933)-Slutzky (1937) paradigm and refer to this as a *shocked potential outcome time series*. We show that assuming that the treatments are shocks enormously simplifies the study of dynamic causal effects. If the treatments are shocks and the causal effects are additive, a wide variety of dynamic causal estimands become equivalent. We emphasize that this assumption plays *no* role in the identification of dynamic causal effects – the identifying assumption remains non-anticipation. It only simplifies analysis by collapsing the differences between several dynamic causal estimands. We believe that the typical focus on shocks is the core reason that much of the existing macroeconometric literature on dynamic causal effects has overlooked the subtle yet important differences between the estimands that we introduce.

Finally, we use the shocked potential outcome time series framework to clarify the causal content of several common methods in macroeconometrics. We consider local projections (LP), which estimate dynamic causal effects by directly regressing an observed treatment on the outcome of interest at a variety of lags (Jordá, 2005), and provide sufficient conditions such that LP identifies a well-defined causal effect. A recent generalization of local projections uses an instrumental variable as a proxy for an unobserved treatment (Jordá et al., 2015; Stock and Watson, 2018). This is referred to as LP-IV and we show that the existing assumptions of LP-IV are not sufficient to identify a dynamic causal effect and provide the extra sufficient condition. We also consider structural vector autoregressions (SVAR) – a foundational tool for causal inference in time series (Sims, 1980) – and place SVARs into the shocked potential outcome time series framework. We show that an SVAR is causally equivalent to a restricted form of the shocked potential outcome time series under the usual invertibility assumption. Taken together, these results illustrate that the shocked potential outcome time series framework provides a natural starting point for causally interpreting existing empirical techniques in macroeconometrics.

Our work does not appear in a vacuum. Over the last several decades, statisticians and applied microeconometricians have made enormous progress by defining causal effects nonparametrically as the comparison of potential outcomes (Imbens and Rubin, 2015). This has been used to explore

the nonparametric causal content of well-established empirical strategies such instrumental variables estimation (Imbens and Angrist, 1994; Angrist et al., 1996; Mogstad et al., 2018). It has also spurred the development of new tools for estimation and inference. Athey and Imbens (2017) and Abadie and Cattaneo (2018) provide recent reviews of this literature. Our work is closely related to the literature on dynamic treatment effects in small-T, large-N panels. This work has been hugely inspiring for us and we are particularly influenced by the groundbreaking work of Robins (1986). The literature on causal inference in panel data is enormous. Canonical references in statistics and microeconometrics include Murphy et al. (2001), Murphy (2003), Abbring and Heckman (2007) and Heckman and Navarro (2007), while recent work includes Heckman et al. (2016), Boruvka et al. (2018), Blackwell and Glynn (2018) and Hernan and Robins (2019). Throughout the paper, we will continue to connect our work with important research on causal inference in cross-sectional and panel settings.

Inference on dynamic causal effects is one of the great themes of the broader time series literature. Researchers quantify causality in time series in a variety of ways such as using "Granger causality" (Wiener, 1956; Granger, 1969), highly structured models such as DSGE models in macroeconomics (Herbst and Schorfheide, 2015), behavioral game theory (Toulis and Parkes, 2016), state space modelling (Harvey and Durbin, 1986; Harvey, 1996; Bondersen et al., 2015), Bayesian structural models (Brodersen et al., 2015) as well as intervention analysis (Box and Tiao, 1975). The closest work to this paper is Angrist and Kuersteiner (2011) and Angrist et al. (2018), which also study time series using potential outcomes (see also White and Lu (2010) and Lu et al. (2017)). That work is importantly different from our own as it avoids discussion of treatment paths, defining potential outcomes as a function of a single prior treatment and also focus only on superpopulation causal effects. In a similar spirit, Kuersteiner et al. (2018) import methods from microeconometrics, extending the regression discontinuity design to estimate impulse response functions in time series settings. However, that work avoids discussion of potential outcomes altogether.

The rest of this paper is structured as follows. In Section 2, we establish our main causal framework by defining a potential outcome time series and dynamic causal effects, as well as discuss important concepts such as additive causal effects. We also introduce a series of causal estimands such as the weighted causal effect, the causal response function and clarify the role of the impulse response function. Section 3 develops a nonparametric estimator for dynamic causal effects, thereby illustrating how the nonparametric potential outcome time series framework can be used to directly develop new estimation and inference strategies. We then transition to linear methods. In Section 4, we introduce the shocked potential outcome time series and discuss its close links to the Frisch (1933)-Slutzky (1937)

macroeconomic paradigm. Section 5 causally interprets the LP and LP-IV methods using the shocked potential outcome time series framework. Section 6 compares and contrasts the local projection and nonparametric approaches on two well-known empirical examples. Section 7 discusses structural vector autoregressions. Section 8 concludes. Unless stated otherwise, we place longer proofs in Appendix A for ease of exposition. We also collect a series of additional results including a Monte Carlo study in our Web Appendix, labelled here Appendix B. Throughout the paper, we use standard path notation. For a time series $\{A_t : t = 1, 2, \dots T\}$, let $A_{1:t} \equiv (A_1, \dots, A_t)$.

2 Potential outcome time series and causal effects

We begin by laying out the potential outcome time series framework that is largely established in Bojinov and Shephard (2019) for experiments on time series and adapt it to the observational settings that are common in macroeconomics. We extend it to multivariate treatments and outcomes as well as specialize the non-anticipating treatment assumption in a manner that can be implemented in observational settings.

2.1 Potential outcome time series

Suppose there is a single unit that is observed over $t = 1, \dots, T$ periods. At each time period, the unit receives a n_w-dimensional vector of treatments W_t and we observe a n_y-dimensional vector of outcomes Y_t. A *potential outcome* is associated with each observed outcome, which describes what would be observed at time t for a particular path of treatments.

Definition 1. *The time-t k-th* ***potential outcome*** *is written as* $Y_{k,t}(w_{1:T})$ *for* $k = 1, 2, ..., n_y$*, where* $w_{1:T} \in \mathcal{W}^T$ *is a treatment path and* $w_t \equiv (w_{1,t}, \dots, w_{n_w,t})'$ *has compact support* $\mathcal{W} \subset \mathbb{R}^{n_w}$*.*

Write $Y_t(w_{1:T}) \equiv (Y_{1,t}(w_{1:T}), \dots, Y_{n_y,t}(w_{1:T}))'$. For each possible treatment path, there is a different vector of potential outcomes.[2]

As initially defined, the potential outcomes can depend on future treatments. It will be useful to restrict the potential outcomes to only depend on the sequence of treatments up to the current period.

[2]In cross-sectional and panel settings, it is familiar to interpret the potential outcomes for a particular unit as non-random. Then uncertainty arises due to the random assignment of treatments. If the potential outcomes themselves are modeled as random variables, this is typically due to random sampling from some population of interest. Abadie et al. (2017) and Abadie et al. (2018) refer to the former view as "design-based" uncertainty and the latter view as "sampling-based" uncertainty. Here we allow for the potential outcome functions to either be non-random or random conditional on the treatments, which will depend on the underlying model for the time series in applications.

Assumption 1 (Non-anticipating potential outcomes)**.** *For each* $t = 1, \ldots, T$, $Y_t(w_{1:t}, w_{t+1:T}) = Y_t(w_{1:t}, w'_{t+1:T})$ *for all* $w_{1:T} \in \mathcal{W}^T, w'_{t+1:T} \in \mathcal{W}^{T-t}$.

With Assumption 1, we drop the potential outcomes dependence on future treatments and write $Y_{k,t}(w_{1:t})$ for $k = 1, \ldots, n_y$ and similarly $Y_t(w_{1:t})$. Think of $Y_t(w_{1:t})$ as the outcomes if the "state of nature" followed the path $w_{1:t}$. The collection of potential outcomes at time t is written as

$$Y_t(\bullet) \equiv \{Y_t(w_{1:t}) : w_{1:t} \in \mathcal{W}^t\}$$

and the potential outcome paths up to time t as $Y_{1:t}(\bullet) \equiv \{Y_1(\bullet), \ldots, Y_t(\bullet)\}$. This describes the set of all possible paths of potential outcomes, while $y_{1:T}(\bullet)$ is the realized version of $Y_{1:T}(\bullet)$.[3]

We next introduce an assumption on the treatment path assignment mechanism of $W_{1:T}$ and assume that it is non-anticipating. That is, we restrict the treatment assignment mechanism to only depend on past outcomes and treatments.

Assumption 2 (Non-anticipating treatment paths)**.** *For each* $t = 1, \ldots, T$ *and all* $w_t \in \mathcal{W}$ *and* $w_{1:t-1}$ *and* $y_{1:T}(\bullet)$*, the treatments obey*

$$\begin{aligned}&\Pr\{W_t \leq w_t | W_{1:t-1} = w_{1:t-1}, Y_{1:t-1}(w_{1:t-1}) = y_{1:t-1}(w_{1:t-1}), Y_{t:T}(w_{1:t-1}, \bullet) = y_{t:T}(w_{1:t-1}, \bullet)\} \\ &= \Pr\{W_t \leq w_t | W_{1:t-1} = w_{1:t-1}, Y_{1:t-1}(w_{1:t-1}) = y_{1:t-1}(w_{1:t-1})\}.\end{aligned}$$

The time-t treatment is conditionally independent from the current and future potential outcomes.[4] In other words, this means that $Y_{t:T}(w_{1:t-1}, \bullet)$ does not Granger cause W_t (Sims, 1972; Chamberlain, 1982; Engle et al., 1983; Kuersteiner, 2010; Lechner, 2011; Hendry, 2017). We interpret Assumption 2 as the time-series analogue of unconfoundedness. It will allow us to develop estimators for a variety

[3] The attraction of using potential outcomes given treatment paths is deep, linking to Savage (1954) subjective probability for each path. They have a storied history in econometrics and statistics (Neyman, 1923; Roy, 1951; Kempthorne, 1955; Cox, 1958; Rubin, 1974; Robins, 1986). Textbook surveys include Angrist and Pischke (2009), Imbens and Rubin (2015) and Hernan and Robins (2019). Robins (1986) pioneered the use of potential outcomes and treatment paths for panel data. Robins et al. (1999) used them for binary time series and Bondersen et al. (2015) used them for state space models. Recently, Bojinov and Shephard (2019) and Blackwell and Glynn (2018) use outcome and treatment paths in more general settings. Similar in spirit to this work, Angrist and Kuersteiner (2011) and Angrist et al. (2018) also study time series using potential outcomes but avoid paths, instead focusing on $Y_{k,t}(w_{t-p})$ for discrete policy treatments w_{t-p}. Dawid (2000) provides a classic critique of potential outcomes.

[4] In panel data settings, Robins (1994), Robins et al. (1999) and Abbring and van den Berg (2003) use this type of "selection on observables" assumption for the treatment paths $W_{1:T}$. Angrist and Kuersteiner (2011) and Angrist et al. (2018) use a related assumption on time series but without treatment paths. Bojinov and Shephard (2019) use the weaker assumption that : $\Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:T}(\bullet)) = \Pr(W_t \leq w_t | W_{1:t-1}, Y_{1:t-1}(\bullet))$. The difference is unimportant, but the one we use here is easier to implement in observational studies. The Bojinov and Shephard (2019) assumption is equivalent to the "latent sequential ignorability" assumption of Ricciardi et al. (2016) when $T = 2$. In a broader context, this type of assumption is called "latent ignorable" (Frangakis and Rubin, 1999).

of dynamic causal estimands of interest that have desirable properties. Aside from Assumption 2, we place no further restrictions on the treatment paths for now.

We finally make an assumption about the data that the econometrician observes. Denote the observed treatment and outcome paths as $w_{1:T}^{obs}$ and $y_{1:T}^{obs}$ respectively.

Assumption 3 (Observed data)**.** *The observed outcome path obeys* $y_{1:T}^{obs} = y_{1:T}(w_{1:T}^{obs})$.

We are now ready to state the basic causal framework for this paper.

Definition 2 (Potential outcome time series)**.** *A time series of treatments and potential outcomes that satisfies Assumptions 1, 2 and 3 is a* ***potential outcome time series****.*

Remark 2.1 (Expectations of future treatments and non-anticipation)**.** *Macroeconomists often consider how treatment decisions today are influenced by the distribution of future outcomes and how they in turn vary with treatments. For example, consumers and firms are modelled as forward-looking and so, expectations about future outcomes influence behavior today. A simple optimization-based version of this (e.g. in the tradition of Muth (1961), Lucas (1972), Sargent (1981)) is:*

$$W_t = \arg\max_{w_t} \mathbb{E}[\max_{w_{t+1:T}} U(Y_{t:T}(w_{1:T}), w_{t:T}) \,|\, Y_{1:t-1}(W_{1:t-1}), W_{1:t-1} = w_{1:t-1}], \tag{1}$$

where U is a utility function of future outcomes and treatments. The expectation is over the law of $Y_{t:T}(\bullet)$. This decision rule delivers $W_t, Y_t(W_{1:t})$. This is a potential outcome time series, where non-anticipation Assumptions 1 and 2 hold.[5] *In Equation (1), the treatment W_t is a deterministic function of past data. The potential outcome time series framework that we laid out is more general.*

Finally, we introduce an example of a dynamic causal effect in macroeconomics and place it in a potential outcome time series to illustrate our core assumptions.

Example 1 (Monetary policy)**.** *Suppose the Federal Reserve raises the Federal funds rate by 25 basis points. What is the dynamic causal effect of this change on unemployment, Y_t? For now, treat the Federal funds rate itself as the treatment W_t. The non-anticipating potential outcomes Assumption 1 restricts unemployment in the current period Y_t to not depend on future realizations of the Federal Funds rate $W_{t+1:T}$ conditional on current and past realizations of $W_{1:t}$. Expectations about future values of the Federal Funds rate may affect the unemployment rate today. The non-anticipating treatment Assumption 2 restricts the choice of the Federal Funds rate at time t to only depend on past*

[5] The non-anticipation assumptions are similarly plausible if a different model for expectations is used. "Natural expectations" as in Fuster et al. (2010) or "diagnostic expectations" as in Bordalo et al. (2018) both only allow current decisions to depend on (possibly biased) beliefs about future outcomes, not the exact realizations along alternative paths $Y_t(w_{1:t})$.

potential outcomes of the unemployment rate $Y_{1:t-1}$ and past choices $W_{1:t-1}$. In observational data, the treatments may not be non-anticipating and Assumption 2 may not hold. For instance, the Federal Reserve may have private information (e.g. information about the health of the financial system) that helps predict future potential outcomes. However, by increasing the dimension of Y_t to include such observable financial information, Assumption 2 could be made to hold. The issue of unobserved omitted variables is discussed more extensively in Section B.1 of the Web Appendix.

2.2 Time-t causal effects

2.2.1 General case

Causal effects are comparisons of potential outcomes at a particular point in time along different treatment paths. We follow the formal definition provided in Bojinov and Shephard (2019). While similar, this is different from Robins (1986) and subsequently Angrist and Kuersteiner (2011), Boruvka et al. (2018) and Angrist et al. (2018). Each focus on a superpopulation version of causal effects that are often referred to as "average causal effects." In contrast, we work in the cross-sectional tradition that manipulates causal effects without reference to superpopulations (Imbens and Rubin, 2015). This will offer new opportunities in the context of time series.

Definition 3 (Causal effect)**.** *For a potential outcome time series, the* ***time-t causal effect of*** *W* ***on*** *Y for treatment path $w_{1:t}$ and counterfactual path $w'_{1:t}$ is*

$$\tau_t(w, w') \equiv Y_t(w_{1:t}) - Y_t(w'_{1:t}).$$

The time-t causal effect on the k-th outcome is $\tau_{k,t}(w, w') \equiv Y_{k,t}(w_{1:t}) - Y_{k,t}(w'_{1:t})$.

Suppose we move the j-th treatment at time $t-p$, $w_{j,t-p}$, and measure the change in the k-th outcome $p \geq 0$ periods later, $Y_{k,t}(w_{1:t})$. Bojinov and Shephard (2019) refer to these as "p-lag causal effects." We consider two versions of the p-lag causal effect.

Definition 4 (k, j-th lag-p time-t causal effects)**.** *For a potential outcome time series and scalars w, w', let $w_{1:t} = (w_{1:t-p-1}, w_{1:j-1,t-p}, w, w_{j+1:n_w,t-p}, w_{t-p+1:t})$ be the treatment path and let $w'_{1:t} = (w_{1:t-p-1}, w_{1:j-1,t-p}, w', w_{j+1:n_w,t-p}, w^*_{t-p+1:t})$ be the counterfactual treatment path. Then, the* ***k, j-th lag-p time-t causal effect*** *is*

$$\tau_{k,j,t}(w, w')(p) \equiv Y_{k,t}(w_{1:t}) - Y_{k,t}(w'_{1:t}).$$

When $w^*_{t-p+1:t} = w_{t-p+1:t}$*, this is the* k,j**-th lag-**p **time-**t **impulse causal effect**, *denoted as* $\tau^*_{k,j,t}(w,w')(p)$.

The impulse causal effect only allows the j-th treatment at time $t-p$ to vary between the treatment and counterfactual paths. The more general case, which was not discussed in Bojinov and Shephard (2019) and is important for macroeconomic applications, allows the treatments to subsequently vary after time $t-p$. Heuristically, we think of the time-t impulse causal effect as analogous to a partial derivative and the general time-t causal effect as a total derivative.

2.2.2 Causal equivalence

To analyze the causal connections between alternative models we define "causal equivalence."

Definition 5 (Causal equivalence)**.** *Assume* $\{Y_{1:T}(\bullet), X_{1:T}(\bullet)\}$ *is a potential outcome time series. If, for every treatment path* $w_{1:t}$ *and counterfactual path* $w'_{1:t}$*, the time-*t *causal effects satisfy*

$$Y_t(w_{1:t}) - Y_t(w'_{1:t}) \overset{as}{=} X_t(w_{1:t}) - X_t(w'_{1:t}),$$

then $Y_{1:t}(\bullet)$ *and* $X_{1:t}(\bullet)$ *are* ***causally equivalent****.*

To understand causal equivalence, vary $w_{1:t}$ solely, keeping the counterfactual fixed at some reference path $\bar{w}_{1:t}$. If $Y_{1:T}(\bullet)$ and $X_{1:T}(\bullet)$ are causally equivalent, then $Y_t(w_{1:t}) = X_t(w_{1:t}) + U_t$ as $w_{1:t}$ varies, where $U_t = Y_t(\bar{w}_{1:t}) - X_t(\bar{w}_{1:t})$ is fixed and does not vary with $w_{1:t}$. We label U_t a "causal nuisance." It may produce different time-series properties in $Y_t(w_{1:t})$ and $X_t(w_{1:t})$ but it does not generate differing time-t causal effects. So, put in another way, if $Y_{1:T}(\bullet), X_{1:T}(\bullet)$ are causally equivalent, then they are equal up to a causal nusiance. Finally, causal equivalence depends on the scale of the outcomes. That is, $Y_t(\bullet)$ and $cY_t(\bullet)$ are not causally equivalent unless the constant $c = 1$.

2.2.3 Additive casual effects

We sometimes specialize the time-t causal effects to the additive case to link our results to the existing macroeconomic literature on dynamic causal effects, which relies heavily on linear models.

Definition 6 (Additive time-t causal effects and I-additivity)**.** *Assume a potential outcome time series. If, for every treatment path* $w_{1:t}$ *and counterfactual path* $w'_{1:t}$*,*

$$\tau_t(w,w') = \sum_{s=0}^{t-1} \beta_{s,t}(w_{t-s} - w'_{t-s}),$$

where $\beta_{s,t}$ are non-stochastic, $n_y \times n_w$-dimensional matrices, then the time-t causal effects are ***additive****. The matrices $\beta_{s,t}$ are the* ***causal coefficients****. The causal effects are* ***time-invariant additive*** *or* ***I-additive*** *if $\beta_{s,t} = \beta_s$ for $t = 1, 2, ..., T$,*

To draw an analogy, additive time-t causal effects are the time series analogue of linear causal effects in cross-sectional settings, $Y_i(t) = \alpha_i + \beta_i t$ for the i-th unit. Similarly, the restriction of I-additivity is akin to the additional restriction that $\alpha_i = \alpha, \beta_i = \beta$ for all units.

Before continuing, we state two immediate results.

Proposition 2.1. *For a potential outcome time series, the following holds:*

i. Under time-t additive causal effects, the k, j-th lag-p time-t causal effect is $\tau_{k,j,t}(w, w')(p) = \beta_{p,k,j,t}(w - w') + \sum_{s=1}^{p} \sum_{i=1}^{n_w} \beta_{s,k,i,t}(w_{i,t-p+s} - w'_{i,t-p+s})$, where $\beta_{p,k,j,t}$ is the k, j-th element of $\beta_{p,t}$. The k, j-th lag-p time-t impulse causal effect is $\tau^*_{k,j,t}(w, w')(p) = \beta_{p,k,j,t}(w - w')$.

ii. Assume $\{Y_{1:T}(\bullet), X_{1:T}(\bullet)\}$ is a potential outcome time series with I-additive causal effects and let $\beta^Y_{0:T-1}$ and $\beta^X_{0:T-1}$ be their respective causal coefficients. Then, $Y_{1:T}(\bullet)$ and $X_{1:T}(\bullet)$ are causally equivalent for $t = 1, 2, ..., T$ if and only if $\beta^Y_{0:T-1} = \beta^X_{0:T-1}$.

2.2.4 Time-t weighted causal effects

We now introduce a series of causal estimands that are built upon the k, j-th lag-p time-t causal effects. When we calculate $\tau_{k,j,t}(w, w')(p) = Y_{k,t}(w_{1:t}) - Y_{k,t}(w'_{1:t})$ either using the impulse form or general form, there are many possible treatments and counterfactual paths that are consistent with passing through $w_{j,t-p} = w$ and $w'_{j,t-p} = w'$. Each possible path leads to a valid time-t causal effect. Bojinov and Shephard (2019) focus on k, j-th lag-p time-t causal effects of the form

$$\tau_{k,j,t}(w, w')(p) = Y_{k,t}(w^{obs}_{1:t-p-1}, w_{t-p:t}) - Y_{k,t}(w^{obs}_{1:t-p-1}, w'_{t-p:t}), \tag{2}$$

and then use freely selectable weights to summarize different possible paths from time $t - p$ to time t.

The structure in Equation (2) uses the observed treatment path $w^{obs}_{1:t-p-1}$ in both the treatment and counterfactual. The price paid for this simplification is that we restrict attention to causal effects along the observed treatment path $w^{obs}_{1:t-p-1}$. The weights are then used to average over the time-t causal effects along different, future treatment paths. The particular choice of the weights is application-specific and chosen to produce the most relevant average for the application. Section B.2 discusses general weight functions. Here we focus on one particular choice that we will focus on: the weighted causal effect.

Definition 7 (Weighted causal effect)**.** *For a potential outcome time series, the* k, j***-th lag-****p ***time-****t ***weighted causal effect*** $\bar{\tau}_{k,j,t}(w, w')(p)$ *is*

$$\mathbb{E}^{W\,|\,Y(\bullet)}[Y_{k,t}(W_{1:t})\,|\,w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p} = w] - \mathbb{E}^{W\,|\,Y(\bullet)}[Y_{k,t}(W_{1:t})\,|\,w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p} = w'].$$

The expectation operator $\mathbb{E}^{W\,|\,Y(\bullet)}[\,\cdot\,]$ *is defined with respect to the law of*

$$W_{-j,t-p:t}\,|\,\{w^{obs}_{1:t-p-1}, y^{obs}_{1:t-p-1}, W_{j,t-p} = w, Y_{t-p:t}(\bullet)\},$$

where $w_{-j,t-p:t} = (w_{1:j-1,t-p}, w_{j+1:n_w,t-p}, w_{t-p+1:t})$ *denotes all elements of* $w_{t-p:t}$ *except* $w_{j,t-p}$*.*

In Definition 7, the averaging is over the random treatment path, conditional on some aspects of the history of the treatments and varying the j-th treatment at time $t-p$. The potential outcomes here are fixed. The *average* k, j*-th lag-*p *weighted causal effect* is simply a time average of the weighted causal effects and is written as

$$\bar{\tau}_{k,j}(w, w')(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} \bar{\tau}_{k,j,t}(w, w')(p).$$

2.2.5 Time-t causal response function

So far we have studied causal effects conditional on $Y_{t-p:t}(\bullet)$. Now view $Y_{t-p:t}(\bullet)$ as random, joining the already random treatment path $W_{t-p:t}$, averaging the weighted causal effect $\bar{\tau}_{k,j,t}(w, w')(p)$ over the potential outcomes $Y_{t-p:t}(\bullet)$ holding fixed the observed outcomes $y^{obs}_{1:t-p-1}$. By applying iterative expectations to the weighted causal effect, we get the following object.

Definition 8 (Causal response function)**.** *For a potential outcome time series, let the expectation operator* $\mathbb{E}^{W,Y}$ *denote the expectation over the joint law of the treatment path* $W_{1:t}$ *and the potential outcomes* $Y_{1:t}(\bullet)$*. Define the* k, j**-th lag-**p **time-**t **causal response function** *as*

$$CRF_{k,j,t}(w, w')(p) \equiv \mathbb{E}^{W,Y(\bullet)}[\bar{\tau}_{k,j,t}(w, w')(p)\,|\,y^{obs}_{1:t-p-1}, w^{obs}_{1:t-p-1}].$$

If the expectations exist, then $CRF_{k,j,t}(w, w')(p)$ *can be written as*

$$\mathbb{E}^{W,Y(\bullet)}[Y_{k,t}\,|\,y^{obs}_{1:t-p-1}, w^{obs}_{1:t-p-1}, W_{j,t-p} = w] - \mathbb{E}^{W,Y(\bullet)}[Y_{k,t}\,|\,y^{obs}_{1:t-p-1}, w^{obs}_{1:t-p-1}, W_{j,t-p} = w'],$$

where $Y_{k,t} = Y_{k,t}(W_{1:t})$ *as shorthand.*

The temporal average

$$CRF_{k,j}(w,w')(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} CRF_{k,j,t}(w,w')(p),$$

is the *average k,j-th lag-p causal response function.* While similar, the causal response function is different than the usual impulse response function defined in the macroeconometric literature as the causal response function conditions on the observed treatment and outcome paths. We also note that the causal response function is similar to the estimand of interest in Angrist et al. (2018).

2.2.6 Is the impulse response function causal?

We now turn to the dominant causal estimand of interest in the macroeconometric literature – the impulse response function (IRF). It is standard to study dynamic causal effects through the IRF.

Definition 9 (Impulse response function)**.** *Assume $Y_{k,t} = Y_{k,t}(W_{1:t})$ and $W_{j,t}$ are jointly strictly stationary and*

$$IRF_{k,j}(w,w')(p) \equiv \mathbb{E}[Y_{k,t} \,|\, W_{j,t-p} = w] - \mathbb{E}[Y_{k,t} \,|\, W_{j,t-p} = w'],$$

exists, where here $\mathbb{E}[\cdot]$ is calculated from the joint law of $Y_{k,t}, W_{j,t-p}$. Then, $IRF_{k,j}(w,w')(p)$ is an ***impulse response function*** *(IRF) as p varies.*

IRFs were introduced by Sims (1980) for vector autoregressions. Reviews of this literature include Kilian (2011), Ramey (2016), Stock and Watson (2016) and Kilian and Lutkepohl (2017). The IRF is widely interpreted causally. As the index p changes, IRFs are viewed as tracing out the dynamic causal effect of the j-th treatment on the k-th outcome (Stock and Watson, 2018).

However, without additional assumptions, the IRF has *no* causal meaning. It is just the difference of two conditional expectations where the conditioning variable is evaluated at two different values. Causal inference is about measuring what happens if $W_{j,t-p}$ is *moved* from w to w'. The IRF does not in general answer that question. This is well known in the literature, which typically works with IRFs in the context of parametrized causal models such as the structural vector moving average in Sims (1980). The additional parametric assumptions gave the IRFs causal content.

We can directly link the IRF to the CRF through the following result. This shows that the IRF has a nonparametric causal meaning if phrased within the potential outcome time series framework and we are able to use strict stationarity.

Theorem 2.1. *Assume a potential outcome time series and* $Y_{k,t} := Y_{k,t}(W_{1:t})$ *and* $W_{j,t}$ *are jointly strictly stationary. Then, if the expectations exist,* $\mathbb{E}^{W_{1:t-p-1},Y_{1:t-p-1}}[CRF_{k,j,t}(p)] = IRF_{k,j}(w,w')(p)$.

Proof. If the expectations exist, then $\mathbb{E}^{W_{1:t-p-1},Y_{1:t-p-1}}[CRF_{k,j,t}(p)] = \mathbb{E}[Y_{k,t}(W_{1:t}) \,|\, W_{j,t-p} = w] - \mathbb{E}[Y_{k,t}(W_{1:t}) \,|\, W_{j,t-p} = w']$, and the RHS is the IRF. □

In Theorem 2.1, the joint distribution of paths and outcomes is used to average over the older history of the treatment path. In this sense, the IRF is attractive as it does not depend upon the time period t, the observed outcomes $y^{obs}_{1:t-p}$ nor the observed treatments $w^{obs}_{1:t-p}$. However, this averaging is a form of extrapolation, assuming the form of causality seen in the sample holds universally over time. This may be true but it is a substantial additional leap to make.

3 Nonparametric estimation of dynamic causal effects

In this section, we discuss nonparametric inference for dynamic causal estimands. Much of the existing macroeconometric work focuses on two linear estimation methods – local projection and structural vector autoregressions. Beginning in Section 4, we will directly discuss these linear approaches.

3.1 Background

Before continuing, we briefly discuss some existing causal work in macroeconomics as motivation. On top of the usual problem of causal inference that it is not possible to observe both $Y_t(W_{1:t})$ and $Y_t(W'_{1:t})$, there is an additional challenge in macroeconomics – the treatments of interest $W_{1:t}$ are hard to measure. For example, a monetary policy shock refers to the "surprise" or "unexpected" component of monetary policy announcements and so, measures of this shock must be constructed from data. The same is true of technology shocks and energy price shocks. These constructed measures are then used as an input to estimating dynamic causal effects. Putting these difficulties aside, it is common to attempt to directly link treatments $W_{j,t-p}$ to outcomes $Y_{k,t}$ in macroeconomics, assuming that the treatment of interest is accurately measured.

One of the earliest attempts to directly measure macroeconomic shocks is Romer and Romer (1989), which uses a "narrative approach", scouring the minutes from Federal Reserve Open Market Committee meetings to identify six post World War II cases in which the Federal Reserve intervened in an attempt to reduce inflation by engineering a recession. The authors then use these historical episodes to estimate the causal effect of monetary policy shocks on macroeconomic variables, $Y_{k,t}$. Following this, Romer and Romer (2004) construct a time series of monetary policy shocks over the

period 1969-1996 using the narrative approach and Romer and Romer (2010) analyze presidential speeches and congressional records to construct measures of fiscal policy shocks. Mertens and Ravn (2013) also construct a series of narratively identified, exogenous federal tax liability changes.

In a similar spirit, Rudebusch (1998) constructs measures of monetary policy shocks by using high-frequency changes in the rates of Federal funds futures contracts in a small window around FOMC announcements. A large literature has since directly regressed a variety of real and financial macroeconomic variables on such constructed shocks to study the causal effect of monetary policy shocks on the macroeconomy (e.g. Kuttner (2001), Cochrane and Piazessi (2002), Faust et al. (2003), Bernanke and Kuttner (2005), Gurkaynak et al. (2005) and recently Nakamura and Steinsson (2018)). Hamilton (2003) and Kilian (2008) construct a measure of oil price shocks to estimate the "causal effect" of oil price shocks on the macroeconomy. More recently, Ramey and Zubairy (2018) use regressions to estimate fiscal multipliers on government spending by using historical information on changes in defense spending to measure fiscal policy shocks.

3.2 Nonparametric estimator of the weighted causal effect

Suppose we observe both the outcome $Y_{k,t} = Y_{k,t}(W_{1:t})$ and treatment of interest $W_{j,t-p}$. This is likely after a great deal of measurement work. For each time period t, we develop an estimator of the k,j-th time-t weighted causal effect $\bar{\tau}_{k,j,t}(w,w')(p)$ in Definition 7 and its temporal average $\bar{\tau}_{k,j}(w,w')(p)$. The properties of the estimator will be derived by regarding the non-anticipating treatment path as random holding the potential outcomes fixed.[6]

We use the triangular filtration notation $\mathcal{F}_{T,t}$ (pg. 53 of Hall and Heyde (1980)) to denote the history $\{w^{obs}_{1:t}, y^{obs}_{1:t}, Y_{t+1:T}(\bullet)\}$. We will also write the weighted causal effect $\bar{\tau}_{k,j,t}(w,w')(p)$ as

$$\mathbb{E}^{W\,|\,Y(\bullet)}_{T,t-p-1}[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})\,|\,W_{j,t-p}=w] - \mathbb{E}^{W\,|\,Y(\bullet)}_{T,t-p-1}[Y_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})\,|\,W_{j,t-p}=w'] \qquad (3)$$

to emphasize that the expectation is taken conditional on this filtration.

In this subsection, we follow the vast bulk of the nonparametric causal literature in assuming that the support of the treatments, $\mathcal{W}$, is discrete. We will deal with the continuous treatment case later in this subsection. To make progress, we additionally assume that the non-anticipating treatment is probabilistic. Denote $p_{j,t}(w) \equiv \Pr(W_{j,t} = w\,|\,\mathcal{F}_{T,t-1})$.

[6] Fixing the potential outcomes follows in the microeconometrics tradition discussed by Imbens and Rubin (2015), Abadie et al. (2017) and Abadie et al. (2018) and traces back to Fisher (1925, 1935) and Cox (1958). Bojinov and Shephard (2019) first introduced this approach into time series experiments.

Assumption 4 (Probabilistic treatment)**.** *For all* $t \geq 1$, $\mathcal{F}_{T,t-1}$ *and* $w \in \mathcal{W}$, $p_{j,t}(w) > 0$.

Assumption 4 is the analogue of the overlap assumption made in cross-sectional settings.[7] We maintain that the treatments are non-anticipating given in Assumption 2.

Define a Horvitz and Thompson (1952) style estimator

$$\hat{\bar{\tau}}_{k,j}(w,w')(p) \equiv \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\bar{\tau}}_{k,j,t}(w,w')(p), \quad \hat{\bar{\tau}}_{k,j,t}(w,w')(p) \equiv \frac{y_{k,t}^{obs}\left\{\mathbb{1}(w_{j,t-p}^{obs}=w) - \mathbb{1}(w_{j,t-p}^{obs}=w')\right\}}{p_{j,t-p}(w_{j,t-p}^{obs})}. \tag{4}$$

This estimator appears in Angrist et al. (2018), but for a different estimand – which is closer to $CRF_{k,j}(w,w')(p)$, the superpopulation estimand that additionally averages over the potential outcomes. Here we focus on the weighted causal effect which fixes the potential outcomes. This type of estimator also appeared in Bojinov and Shephard (2019). Here it is much simpler due to the choice of the weight function which appears in the expectations in the definition of the weighted causal effect.

3.2.1 Properties of $\hat{\bar{\tau}}_{k,j,t}(w,w')(p)$ and $\hat{\bar{\tau}}_{k,j}(w,w')(p)$

The following theorem shows that $\hat{\bar{\tau}}_{k,j,t}(w,w')(p) - \bar{\tau}_{k,j,t}(w,w')(p)$ has martingale difference errors and hence $\hat{\bar{\tau}}_{k,j}(w,w')(p)$ is unbiased, consistent and asymptotically normal under minimal conditions.

Theorem 3.1 (Randomization properties of $\hat{\bar{\tau}}_{k,j,t}(w,w')(p)$)**.** *Assume a potential outcome time series and Assumption 4. Let* $u_{k,j,t-p} := \hat{\bar{\tau}}_{k,j,t}(w,w')(p) - \bar{\tau}_{k,j,t}(w,w')(p)$. *Then, over the law of the non-anticipating treatment path,*

$$\mathbb{E}_{T,t-p-1}^{W\,|\,Y(\bullet)}[u_{k,j,t-p}\,|\,\mathcal{F}_{T,t-p-1}] = 0, \quad \textit{and} \quad \mathbb{E}^{W\,|\,Y(\bullet)}[\hat{\bar{\tau}}_{k,j}(w,w')(p)] = \bar{\tau}_{k,j}(w,w')(p). \tag{5}$$

Further $\eta_{k,j,t-p}^2 \equiv Var^{W|Y(\bullet)}[u_{k,j,t-p}|\mathcal{F}_{T,t-p-1}]$, *exists. If the non-anticipating treatment paths satisfy* $\lim_{T\to\infty}\frac{1}{T-p}\sum_{t=p+1}^{T}\mathbb{E}^{W\,|\,Y(\bullet)}[\eta_{k,j,t-p}^2] < \infty$, *then* $\hat{\bar{\tau}}_{k,j}(p) - \bar{\tau}_{k,j}(p) \xrightarrow{p} 0$ *as* $T \to \infty$. *Finally, if* $\frac{1}{T-p}\sum_{t=p+1}^{T}\bar{\eta}_{k,j,t-p}^2 \xrightarrow{p} \eta_{k,j}^2 > 0$, *then, over the non-anticipating treatment path,*

$$\sqrt{T}\frac{\{\hat{\bar{\tau}}_{k,j}(p) - \bar{\tau}_{k,j}(p)\}}{\eta_{k,j}} \xrightarrow{d} N(0,1). \tag{6}$$

[7] There is a large literature in cross-sectional causal inference that studies the overlap assumption. For example, D'Amour et al. (2018) considers the implications of overlap in high dimensional settings, Busso et al. (2014) analyzes the finite sample performance of estimators in the presence of limited overlap and finally, Khan and Tamer (2010) discuss how the failure of overlap leads to "irregular identification." Although we do not explore them here, these issues extend to the potential outcome time series setting.

Again, the only source of randomness here is the path of the treatments and so Theorem 3.1 describes properties of the randomization distribution of the estimator $\hat{\bar{\tau}}_{k,j}(p)$.

Remark 3.1. *Assume a potential outcome time series and Assumption 4. Then, note:*

1. *The estimation error $u_{k,j,t-p}$ is a triangular martingale difference with respect to the filtration $\mathcal{F}_{T,t-p-1}$. The conditional variance $\eta^2_{k,j,t-p}$ in Theorem 3.1 must exist as we have conditioned on the potential outcomes, the treatments are defined over a compact set and $p_{j,t-p}(w) > 0$ by Assumption 4. The proof of Theorem 3.1 shows $\eta^2_{k,j,t-p}$ is*

$$\mathbb{E}^{W\,|\,Y(\bullet)}_{T,t-p-1}\left(\frac{y^2_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})}{p_{j,t-p}(w)}\,|\,W_{j,t-p}=w\right)+\mathbb{E}^{W\,|\,Y(\bullet)}_{T,t-p-1}\left(\frac{y^2_{k,t}(w^{obs}_{1:t-p-1}, W_{t-p:t})}{p_{j,t-p}(w')}\,|\,W_{j,t-p}=w'\right)-\bar{\tau}^2_{j,k,t}(p).$$

2. *In general, we cannot compute $\eta^2_{k,j,t-p}$ as some of the potential outcomes are unobserved. However, $\dot{\eta}^2_{k,j,t-p} = \eta^2_{k,j,t-p} + \bar{\tau}^2_{j,k,t}(p)$ is an upper bound which can be estimated. It is likely to be quite good unless the causal effect is large. Then, an unbiased estimator of $\dot{\eta}^2_{k,j,t-p}$ is*

$$\hat{\dot{\eta}}^2_{k,j,t-p}(w,w')(p) = \frac{(y^{obs}_{k,t})^2\left\{\mathbb{1}(w^{obs}_{j,t-p}=w)+\mathbb{1}(w^{obs}_{j,t-p}=w')\right\}}{p_{j,t-p}(w^{obs}_{j,t-p})^2}.$$

3. *In practice, proxy outcomes can be used to reduce the variance of $\hat{\bar{\tau}}_{k,j,t}(w,w')(p)$. In cross-sectional work, this approach leads to doubly robust estimators (e.g. Raz (1990), Robins et al. (1994), Rosenbaum (2002), Bang and Robins (2005), Hennessy et al. (2016)). Here we find a proxy that approximates $Y_{k,t}$ given information available at time $t-p-1$. For example, we could use $y^{obs}_{k,t-p-1}$ and so the estimator becomes*

$$\hat{\bar{\tau}}_{k,j,t}(p) = \frac{(y^{obs}_{k,t}-y^{obs}_{k,t-p-1})}{p_{j,t-p}(w^{obs}_{j,t-p})}\left\{\mathbb{1}(w^{obs}_{j,t-p}=w)-\mathbb{1}(w^{obs}_{j,t-p}=w')\right\}.$$

   Again, this is unbiased under the treatment assignment process. In this example, an unbiased estimator of the upper bound of the variance of the proxy is therefore

$$\frac{(y^{obs}_{k,t}-y^{obs}_{k,t-p-1})^2}{p_{j,t-p}(w^{obs}_{j,t-p})^2}\left\{\mathbb{1}(w^{obs}_{j,t-p}=w)+\mathbb{1}(w^{obs}_{j,t-p}=w')\right\}.$$

4. *In Section B.3 of the Web Appendix, we extend the results in Theorem 3.1 to provide an unbiased, nonparametric estimator of the causal response function defined in Definition 8.*

3.2.2 Testing for dynamic causal effects

One way of assessing the significance of the treatment on the outcome is through a sharp null hypothesis (Fisher (1925, 1935), Bojinov and Shephard (2019)) of no temporal causal effect

$$H_0 : Y_{k,t}(w_{1:t}) = Y_{k,t}(w'_{1:t}), \quad \text{for all} \quad w_{1:t}, w'_{1:t}, \quad t = 1, ..., T, \quad k = 1, ..., n_w.$$

Under this sharp null, $y^{obs}_{1:T} = y_{1:T}(w_{1:T})$ for all $w_{1:T} \in \mathcal{W}^T$. Therefore, under the sharp null, all potential outcome paths are $y_{1:T}(\bullet)$ known and so, it is possible to simulate the randomization distribution of an estimator by simulating new treatment paths. New treatment paths can be simulated using the prediction decomposition and the repeated application of Assumption 4 if the treatment probability $p_{j,t}(w)$ is known.[8] These simulations are used to compute the exact null distribution of any causal test statistic against a portmanteaux alternative. In particular, we can compute the exact p-value of the observed statistic and construct exact confidence intervals through test inversion.[9]

The sharp null rules out temporal causal effects at every time period (but see Wu and Ding (2018)). A weaker version is the null hypothesis of no average temporal causal effect at lag-p of the j-th treatment on the k-th outcome as we move the treatment from w to w'. This is written as

$$H_0 : \bar{\tau}_{k,j}(w, w')(p) = 0.$$

In cross-sectional work, this is often called a Neyman null (e.g. Ch. 6 of Imbens and Rubin (2015)).

We can test this null using the central limit theorem given in Theorem 3.1. Under the null H_0, $\frac{\sqrt{T}\hat{\bar{\tau}}_{k,j}(w,w')(p)}{\bar{\eta}_{k,j}(w,w')(p)} \xrightarrow{d} N(0,1)$ as $T \to \infty$. This can be used to form a conservative test whose limiting distribution can still be compared to a standard normal random variable. In particular, consider

$$Z_{k,j}(w, w')(p) = \frac{\sqrt{T}\hat{\bar{\tau}}_{k,j}(w, w')(p)}{\hat{\bar{\eta}}_{k,j}}, \quad \hat{\bar{\eta}}^2_{k,j} = \frac{1}{T-p}\sum_{t=p+1}^{T} \hat{\bar{\eta}}^2_{k,j,t-p}.$$

[8] If $p_{j,t}(w)$ is only known up to a finite number of parameters, $p_{j,t}(w \,|\, \lambda)$, then posterior predictive P-values (Meng (1994)) can be exactly computed by also sampling (under the null) from the posterior $\lambda \,|\, Y_{1:T}(\bullet), W_{1:T}$ (based on the conditional likelihood $W_{1:T} \,|\, \lambda, Y_{1:T}(\bullet)$ and a prior for λ) which gives posterior samples from $p_{j,1:T}(w \,|\, \lambda)$. This allows us to simulate from $W_{1:T} \,|\, Y_{1:T}(\bullet)$ which determines the test statistic and so the posterior predictive P-value of the observed test statistic. Notice this approach requires no modelling of the outcomes, only the j-th treatments. Alternative classical inference based on parametrized models and nonparametric methods is discussed by, for example, Hirano et al. (2003).

[9] Extending this argument, suppose outcomes and treatments are univariate. Then define the lag-p additive hypothesis:

$$H^p_0 : Y_t(w_{1:t}) = Y_t(w'_{1:t}) + \beta_p(w_{t-p} - w'_{t-p}), \quad \text{for all} \quad w_{1:t}, w'_{1:t}, \quad t = 1, ..., T.$$

Fix β_p and compute $y_t(w_{1:t}) = y^{obs}_t + \beta_p(w_{t-p} - w^{obs}_{t-p})$ for all $w_{1:t}$ and t. Then, we simulate from $W_{1:T} \,|\, \{Y_{1:T}(\bullet), \beta_p\}$ to evaluate the p-value of any test statistic given β_p. Exact confidence regions for β_p can be constructed via test inversion.

We reject the null if $Z_{k,j}(w, w')(p)$ is significantly far from 0. This testing approach is nonparametric and only assumes that the non-anticipating treatments are probabilistic. Much progress has been made in the microeconometric literature by focusing attention on tests in this spirit even though they are conservative and asymptotic.[10]

From a more fundamental methodological perspective, note that this test is conditional on the potential outcomes. This is appropriate for testing for causal effects during the observed period of time. Continuing an earlier example, it is reasonable to treat the potential outcomes as fixed when testing the causal effect of changes in the Federal funds rate on unemployment during the 1980s. However, this test provides no guarantees about the causal effects outside of the observed stretch of data. The advantages and disadvantages of building econometric methods that do not attempt to extrapolate beyond the current data is much discussed in the microeconometrics literature (e.g. Angrist and Pischke (2010)). This strict non-extrapolation approach may be less appealing in macroeconometrics (see the discussion in Sims (2010)) as replication of results with new datasets and summarizing the results through meta-analysis is potentially more difficult in many applications of interest. However, non-extrapolation of causal results may well make our causal statements more realistic. Using these base results, we can also derive martingale difference based results for the causal response function. The extension to that case is somewhat routine and is discussed in Section B.3.

3.2.3 Nonparametric estimator with continuous treatments

There is a modest literature on the nonparametric estimation of causal effects when treatments are continuous in cross-sectional and panel settings. For example, Hirano and Imbens (2004) study continuous treatments using "generalized propensity scores." Marginal structural models of Robins et al. (2000) provide parametric and series based nonparametric strategies to deal with continuous treatments. Cattaneo (2010) provides an extensive discussion of the multivalued case and the related literature. Yang et al. (2016) is a recent paper on this topic.

Write $F_{j,t}(w) = \Pr(W_{j,t} \leq w | \mathcal{F}_{T,t-1})$. Then, using a bandwidth $h > 0$, define the time-t estimator

$$\hat{\bar{\tau}}_{k,j,t}(w, w') \equiv \frac{y_{k,t}^{obs} \left(\mathbb{1}(w_{j,t-p}^{obs} \in [w-h, w+h]) - \mathbb{1}(w_{j,t-p}^{obs} \in [w'-h, w'+h]) \right)}{F_{j,t-p}(w_{j,t-p}^{obs} + h) - F_{j,t-p}(w_{j,t-p}^{obs} - h)}. \tag{7}$$

Its temporal average is $\hat{\bar{\tau}}_{k,j}(w, w')(p) := \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{\bar{\tau}}_{k,j,t}(w, w')(p)$. This binned estimator is simply

[10] In Monte Carlo experiments, Bojinov and Shephard (2019) suggest that the distortion produced by the conservative test is modest. This is encouraging evidence but will not hold universally. For example, the distortions may be sizable if the true causal effects are very large.

the difference of two kernel type estimator of the causal effect (e.g. Silverman (1986)). The vital point is that the nonparametrics is only in the single dimension of $W_{j,t-p}$, so there is no "curse of dimensionality" in this problem.

More formally, the optimal mean square error nonparametrics will typically reduce the rate of convergence of $\hat{\bar{\tau}}_{k,j}(w,w')(p)$ from T^{-1} in the discrete case to $T^{-4/5}$ for the continuous case. However, this rate reduction holds uniformly across all of p, n_y and n_w. We formalize this statement in Section B.4 in the Web Appendix. We compute the estimator's bias, variance and the optimal selection of h to optimize the mean square error under various smoothness assumptions.[11] All of this work is routine from the kernel literature. The results from a Monte Carlo experiment grounded in our empirical results from Section 6 are given in Web Appendix Section B.6.

4 Frisch-Slutzky paradigm, shocks and SVMA

4.1 Background

In a potential outcome time series, the non-anticipating treatment path causally affects an outcome path. In the literature on the Frisch (1933)-Slutzky (1937) paradigm in macroeconomics, we do not see non-anticipation type assumptions. Instead, each treatment is assumed to be *unpredictable* given the past, which is a significantly stronger assumption. We label such treatments "shocks." For example, typical shocks of interest are monetary policy shocks, oil price shocks or technology shocks. Here we bring together the literature on shocks with the potential outcome time series.

4.2 Shocked potential outcome time series

When we map shocks to the nonparametric causal framework laid out in Section 2, we view a shock as a particular type of non-anticipating treatment. Recall that under Assumption 2, non-anticipating treatment paths are determined by the law of $W_t \,|\, W_{1:t-1}, Y_{1:t-1}$. We say the time-$t$ treatment is a *shock* if it is non-anticipating and additionally if it is a martingale difference with respect to the filtration generated by the potential outcome time series (e.g. Hall and Heyde (1980)).

[11] A simple way to select the bandwidth is to minimize the sum of the squared bias plus conservative variance: $h_k(p)^4 b_k^2(p) + h_k(p)^{-1} \frac{1}{T-p} v_k^2(p)$, where the bias is, under linearity, $b_k^2(p) = \beta_{p,k}^2 \frac{1}{9(T-p)} \sum_{t=p+1}^{T} \left(d \log f_{t-p}(w)/dw - d \log f_{t-p}(w')/dw \right)^2$. So, the approximate sum of squared bias is large if the density of the treatment is varying strongly and if the squared linear projection treatment effects are large. Hence a reasonable choice is $h_k(p) = (T-p)^{-1/5} \left(0.25 v_k^2(p) b_k^{-2}(p) \right)^{1/5}$. Therefore the bandwidth will be wide if the variance is large compared to the sum of squared bias. We have used this approach in our empirical work discussed in Section 6 and Monte Carlo studies in the Web Appendix Section B.6.

Definition 10 (Treatments as shocks)**.** *For a potential outcome time series, if, for all* $W_{1:t-1}, Y_{1:t-1}$*,*

$$\mathbb{E}[W_t \,|\, W_{1:t-1}, Y_{1:t-1}] = 0, \quad \textit{and} \quad Var(W_t \,|\, W_{1:t-1}, Y_{1:t-1}) = diag\{\sigma^2_{1,t}, \ldots, \sigma^2_{n_w,t}\},$$

then the time-t treatment is a ***shock****.*

This provides the formal definition of a shock in a nonparametric causal framework. First, for a potential outcome time series W_t is *non-anticipating* as defined in Assumption 2. Second, the shock W_t is *unpredictable*. This is often described heuristically in the literature. For example, Stock and Watson (2018) describe macroeconomic shocks as "unanticipated structural disturbances" that produce "unexpected changes" in the macroeconomic outcomes of interest.[12] We additionally assume the elements of the shock are conditionally *mutually uncorrelated*. This is also a standard interpretation of a shock in the macroeconomic literature (Stock and Watson, 2016; Ramey, 2016). Note that the individual conditional variances are free to vary through time which is useful as time-varying volatility is important in macroeconomics (e.g. Stock and Watson (2003) and Justiniano and Primiceri (2008)).

We now bring this together to give a modern causal statement of the Frisch-Slutzky paradigm.

Definition 11 (Shocked potential outcome time series)**.** *Assume a potential outcome time series has treatments which are shocks following Definition 10. Then the system is a* ***shocked potential outcome time series****.*

To make this concrete, we now illustrate the definition of a shock.

Example 1 (continued)**.** *Returning to the monetary policy example, now interpret* W_t *to be a monetary policy shock associated with the Federal Funds rate as defined in Assumption 10.* W_t *is a monetary policy shock if it is unforecastable given the available information up to time* $t-1$*,* $\{W_{1:t-1}, Y_{1:t-1}\}$*, which may include past values of the Federal Funds rate and a long vector of observable macroeconomic indicators. In other words, the monetary policy shock* W_t *is the "surprise" change in monetary policy.*

Now return to the causal effects and the associated causal estimands defined in Section 2 and consider a shocked potential outcome time series. These objects are enormously simplified under additive causality and the martingale difference assumption on the treatments. This is summarized in the following theorem.

[12] Ramey (2016) also describes shocks as: (1) "exogenous with respect to the other current and lagged endogenous variables," (2) "uncorrelated with other exogenous shocks" and (3) "either unanticipated movements in exogenous variables or news about future movements in exogenous variables."

Theorem 4.1. *If a shocked potential outcome time series has additive causal effects, then*

$$CRF_{k,j,t}(p) = \bar{\tau}_{k,j,t}(w, w')(p) = \tau^*_{k,j,t}(w, w')(p) = \beta_{p,k,j,t}(w - w'),$$

for each $t \geq p + 1$. *If additionally* $\{Y_{k,t}(W_{1:t}), W_{j,t}\}$ *is a jointly stationary stochastic process and the causal effects are I-additive, then also*

$$IRF_{k,j}(w, w')(p) = \beta_{p,k,j}(w - w').$$

Proof. Given in the Appendix. □

This is one of our key results. If the treatments are shocks and the causal effects are additive, then the weighted causal effect in Definition 7 and causal response function in Definition 8 are equal. So, it is enormously simplifying to restrict causal effects to be additive and treatments to be shocks. In particular, the choice of causal estimand no longer matters as they all identify the same objects – the causal coefficients. The core causal insights in Theorem 4.1 are driven by the properties of the treatments and the additive causal effects, not the time series properties of the outcomes. If we strengthen the assumptions on the observed outcomes and treatments and invoke stationarity, then the impulse response function in Definition 9 also identifies the causal coefficients. In macroeconometric research, the focus is typically on linear models with shocks and consequently on $\beta_{p,k,j,t}$, the causal coefficients. As a result, this will be our focus for the rest of the paper as we relate our causal framework to common macroeconometric techniques.

4.3 Structural vector moving average form

We now turn to linear versions of the shocked potential outcome time series model. This will connect to much of the existing work on dynamic causal effects in macroeconomics.

Suppose $Y_t(w_{1:t})$ is differentiable with respect to $w_{1:t}$. Then, Taylor expanding about some fixed $\bar{w}_{1:t}$ and writing the $n_y \times n_w$ matrix $\Theta_{s,t} = \partial Y_t(\bar{w}_{1:t})/\partial w'_s$ produces the approximation of $Y_t(w_{1:t})$ as $\hat{Y}_t(w_{1:t}) = Y_t(\bar{w}_{1:t}) + \sum_{s=0}^{t-1} \Theta_{s,t}(w_s - \bar{w}_s) = U_t^Y + \sum_{s=0}^{t-1} \Theta_{s,t} w_s$. Here U_t^Y is a "causal nuisance", as defined in Section 2.2.2, so it does not impact the time-t causal effects. Note that $\hat{Y}_t(w_{1:t}) - \hat{Y}_t(w'_{1:t}) = \sum_{s=0}^{t-1} \Theta_{s,t}(w_{t-s} - w'_{t-s})$ so $\tau^*_{k,j,t}(w, w')(p) = \Theta_{p,k,j,t}$, the k, j-th element of $\Theta_{p,t}$. In general, $\Theta_{s,t}$ will be stochastic with $\mathbb{E}^{W \,|\, Y(\bullet)}[\Theta_{s,t} W_s \,|\, W_{1:s-1}, Y_{1:s-1}] = 0$, which follows as $\bar{w}_{1:t}$ is fixed and $\mathbb{E}^{W|Y(\bullet)}[W_t \,|\, W_{1:t-1}, Y_{1:t-1}] = 0$ since W_t is a shock. In many time series models, $Y_t(w_{1:t})$ is non-

stochastic and so the treatment path $W_{1:t}$ solely drives the stochastic behaviour of $Y_t(W_{1:t})$. In this case, $\Theta_{s,t}$ is non-stochastic for all s, t.

A conceptually different approach is to directly assume that the shocked potential outcome time series has time-t additive causal effects. With this, we can again write

$$Y_t(w_{1:t}) = U_t^Y + \sum_{s=0}^{t-1} \Theta_{s,t} w_{t-s},$$

where now $\Theta_{s,t}$ is a $n_y \times n_w$ fixed matrix and U_t^Y is again a causal nuisance. Strengthen this to I-additivity and this simplifies as we can drop the time subscripts on the matrices, $\Theta_{s,t} = \Theta_s$.

Much of the macroeconometrics literature goes much further than a priori imposing I-additivity and instead directly works with the structural vector moving average (SVMA) (e.g. Sims (1980), Ramey (2016), Kilian and Lutkepohl (2017), Stock and Watson (2016), Stock and Watson (2018)).

Definition 12. *Assume a shocked potential outcome time series and let $\Theta_0, ..., \Theta_{t-1}$ be non-stochastic matrices. If $n_w = n_y$ and for each time period t, $Y_t(w_{1:t}) = \mu + \sum_{s=0}^{t-1} \Theta_s w_{t-s}$ for all $w_{1:t} \in \mathcal{W}^t$, then $Y_{1:t}(\bullet)$ is a structural vector moving average.*

Alternatively, we could have worked with $Y_t(w_{1:t}) = \mu + \sum_{s=0}^{\infty} \Theta_s w_{t-s}$. Note that in the SVMA, the potential outcomes $Y_t(w_{1:t})$ are non-random given the treatment. This is a canonical example in macroeconomics in which the potential outcomes are a deterministic function of the treatment. Note that while we take this as the starting point of analysis, the SVMA is typically motivated in macroeconometrics as arising from an underlying structural model.

We now collect two useful results about the potential SVMA form.

Proposition 4.1. *The following statements hold:*

i. The SVMA has I-additive causal effects with $\tau_t(w, w') = \sum_{s=0}^{t-1} \Theta_s (w_{t-s} - w'_{t-s})$.

ii. The SVMA is causally equivalent to a potential outcome time series with I-additive causal effects.

Proof. Claims (i)-(ii) are immediate from the definitions provided in Section 2 and Proposition 2.1. □

5 Local projection

Here we use the shocked potential outcome time series to provide a common causal foundation to two common linear strategies in macroeconomics: local projection and local projection-IV. We begin

by briefly reviewing each method and then we causally interpret them using the shocked potential outcome time series model.

5.1 Local projection with observed treatments

Assume that, after a great deal of measurement work, we observe both $Y_{k,t} = Y_{k,t}(W_{1:t})$ and $W_{j,t-p}$. A natural strategy is to then directly regress the observed outcomes on the observed treatments at a variety of lags. This strategy is referred to as "local projections" in econometrics (Jordá, 2005).[13] Local projections are extremely common and have been highly influential in macroeconomics. The empirical papers in Section 3.1 take the constructed shocks and then study the causal impacts of the measured treatment $W_{j,k-p}$ using local projections.

5.1.1 LP under additive causality

Our next theorem provides a formal causal interpretation of the local projection method.

Theorem 5.1 (Local projection)**.** *Assume a shocked potential outcome times series as in Definition 11, that $p \geq 0$ and additionally that:*

A. The causal effects are additive, given in Definition 6.

B. $\mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t})] = \alpha$, a time-invariant expectation, where $\bar{w}_{1:t} = 0$ is a fixed known treatment path.

C. $Var^W(W_t) < \infty$ and $Var^Y(Y_{k,t}(\bar{w}_{1:t})) < \infty$ for all t.

Then, for $p \geq 0$, the time-t causal regression of $Y_{k,t}$ on $W_{j,t-p}$ can be written as

$$Y_{k,t}(W_{1:t}) = \alpha + \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}, \quad t = p+1, 2, ..., T, \tag{8}$$

where $\beta_{p,k,j,t}$ is the causal coefficient introduced in Definition 6. Moreover, $\mathbb{E}^{W,Y}[\eta_{k,j,t}] = 0$ and $Cov^{W,Y}(W_{j,t-p}, \eta_{k,j,t}) = 0$ if they exist.

Remark 5.1. *From Theorem 5.1, we have the following notes:*

1. *The error $\eta_{k,j,t}$ can depend upon all of $W_{1:t}$ except for $W_{j,t-p}$ and its dependence on $W_{1:t}$ is made explicit in the proof. Moreover, it is not in general a martingale difference nor even white noise.*

[13] This is related to, but different from, the literature on direct forecasting, which forecasts Y_t by regressing on Y_{t-p} rather than iterating one step ahead forecasts p times (Cox, 1961; Marcellino et al., 2006).

2. *If we strengthen Assumption A to I-additivity, then Equation (8) becomes*

$$Y_{k,t}(W_{1:t}) = \alpha + \beta_{p,k,j} W_{j,t-p} + \eta_{k,j,t}. \tag{9}$$

A natural estimator of the causal coefficient $\beta_{p,k,j}$ is the ordinary least squares estimator

$$\hat{\beta}_{p,k,j} = \frac{\sum_{t=p+1}^{T}(Y_{k,t} - \bar{Y}_k)(W_{j,t-p} - \bar{W}_j)}{\sum_{t=p+1}^{T}(W_{j,t-p} - \bar{W}_j)^2} = \beta_{p,k,j} + \frac{\frac{1}{T-p}\sum_{t=p+1}^{T}\eta_{k,j,t}(W_{j,t-p} - \bar{W}_j)}{\frac{1}{T-p}\sum_{t=p+1}^{T}(W_{j,t-p} - \bar{W}_j)^2}. \tag{10}$$

The uncorrelatedness of $\eta_{k,j,t}$ and $W_{j,t-p}$ drives $\hat{\beta}_{p,k,j}$ to be centered on $\beta_{p,k,j}$. Consistency and a central limit theory can be developed using standard, asymptotic techniques although additional regularity conditions will need to be introduced. Because $\eta_{k,j,t}$ may be serially correlated, HAC/HAR inference techniques may be necessary (Newey and West, 1987; Kiefer et al., 2002; Lazarus et al., 2018). Note that the assumptions of Theorem 5.1 do not imply that $\mathbb{E}^{W,Y}[\eta_{k,j,t} \,|\, W_{j,t-p}] = 0$ and so, the OLS estimator will not in general be unbiased.

3. *A possibly more efficient estimator can be found if there is a control variable $X_{k,j,t}$ that is correlated with the error $\eta_{k,j,t}$ but unrelated to the treatment $W_{j,t-p}$. Then, we fit the model $Y_{k,t}(W_{1:t}) = \alpha^* + \beta_{p,k,j} W_{j,t-p} + \gamma'_{p,k,j} X_{k,j,t} + \eta^*_{k,j,t}$. Here $\gamma_{p,k,j}$ is a nuisance parameter.*

Theorem 5.1 states that if the researcher is willing to assume that the causal effects are additive and that the treatment path is a sequence of shocks, then the local projection method provides information about the causal coefficient $\beta_{p,k,j,t}$ under some mild side conditions. Moreover, we showed in Section 4 that when the treatments are shocks, $\beta_{p,k,j,t}$ is the coefficient of the weighted causal effect, the causal response function as well as the impulse response function under the assumptions of Theorem 5.1. So, if these assumptions hold, the local projection method provides information on each of these objects. In practice, researchers typically work with Equation (9) rather than Equation (8) and so, they implicitly assume I-additive causal effects.

Remark 5.2. *What does the OLS estimator identify if the causal effects are only additive but not I-additive? In Proposition B.5 of our Web Appendix, we show that if the causal effects are additive but not I-additive, then the OLS estimator identifies a time average of the k, j-th lag-p time-t causal coefficients. Additionally, if the causal effects are not additive, we show in Proposition B.6 and Proposition B.7 that the LP estimand can be interpreted as the best linear approximation to the general non-linear time-t causal effects.*

5.2 Local projection with instrumental variables

5.2.1 Background on measurement error

In local projections, we assume that $W_{j,t-p}$ is observed. There is a concern that the economist's measurement $\hat{W}_{j,t-p}$ is in practice a noisy proxy for $W_{j,t-p}$ rather than $W_{j,t-p}$ itself. We now discuss a recent econometric literature that treats the proxy as an instrument for $W_{j,t-p}$. This is referred to as local projection with an instrumental variable (LP-IV) and proceeds in the measurement error spirit of Durbin (1954) and Sargan (1958). Here we provide novel conditions under which this method works. Before we do this, we take a step back and discuss some of the recent literature on this topic.

Since its initial development in Jordá et al. (2015), LP-IV has received much attention. For example, Ramey and Zubairy (2018) estimate fiscal multipliers on government spending and Fieldhouse et al. (2018) estimate the effects of historical credit shocks on asset prices and investment, both using LP-IV methods on constructed measures of underlying shocks. Mertens and Olea (2018) use narratively identified changes in the top marginal tax rate as an instrument in an LP-IV framework to measure the dynamic causal effects on reported income, real GDP and unemployment and Stock and Watson (2018) illustrate LP-IV using constructed high-frequency movements in the rates on federal funds contracts around policy announcements to study the effects of monetary policy shocks. On top of these empirical applications, there has been a recent surge in econometric work that focuses on LP-IV. For example, Stock and Watson (2018) and Plagborg-Møller and Wolf (2018) study LP-IV econometrically and consider the conditions under which the method identifies the usual macroeconometric objects of interest such as the impulse response function of the SVMA form.

5.2.2 Potential outcome time series with added proxies

To use a proxy $\hat{W}_{1:T}$ to make causal statements, we first integrate $\hat{W}_{1:T}$ into our nonparametric potential outcome time series framework.[14] This is necessary to rule out that the proxy $\hat{W}_{1:T}$ drives future potential outcomes, otherwise the causal interpretation will be lost.

Definition 13. *A **proxy potential outcome time series** obeys, for each t and all $w_{1:t}, \hat{w}_{1:T}, \hat{w}'_{1:T}$:*

A. The potential outcomes satisfy $Y_{1:t}(w_{1:t}, \hat{w}_{1:T}) = Y_{1:t}(w_{1:t}, \hat{w}'_{1:T})$. Subsequently we write $Y_{1:t}(w_{1:t})$ and $Y_{1:t}(\bullet)$ as before.

[14] While similar to the approach taken in Imbens and Angrist (1994) and Angrist et al. (1996) on cross-sectional IV, this is importantly different as we do not define a potential treatment function. We proceed in this manner as the instrument is a *proxy* for the treatment in these settings, not an encouragement.

B. The non-anticipating treatment assumption is

$$\begin{aligned}&\Pr\{W_t \leq w_t \,|\, W_{1:t-1} = w_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}(w_{1:t-1}), Y_{t:T}(w_{1:t-1}, \bullet)\} \\ &= \Pr\{W_t \leq w_t \,|\, W_{1:t-1} = w_{1:t-1}, Y_{1:t-1}(w_{1:t-1})\}.\end{aligned} \tag{11}$$

C. Measurement $\hat{W}_t$ satisfies

$$\begin{aligned}&\Pr\{\hat{W}_t \leq \hat{w}_t \,|\, W_{1:t} = w_{1:t}, \hat{W}_{1:t-1}, Y_{1:t-1}(w_{1:t-1}), Y_{t:T}(w_{1:t-1}, \bullet)\} \\ &= \Pr\{\hat{W}_t \leq \hat{w}_t \,|\, W_{1:t} = w_{1:t}, \hat{W}_{1:t-1}, Y_{1:t-1}(w_{1:t-1})\}.\end{aligned} \tag{12}$$

D. Finally: The equality $y^{obs}_{1:T} = y_{1:T}(w^{obs}_{1:T})$ holds and we observe $y^{obs}_{1:T}$ and $\hat{w}^{obs}_{1:T}$.

Assumption A excludes that the proxy for the treatment directly effects outcomes. Assumption B enforces that the proxy provides no additional information on the next treatment beyond past treatments and past potential outcomes. Finally, Assumption C says the proxy can be influenced by the current treatment as well as past information.

With this in hand, we state a lemma that will be helpful later.

Lemma 5.1. *Assume a proxy potential outcome time series. If $W_t, \hat{W}_t$ is a joint martingale difference with respect to $W_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}$, then, if they exist, the superpopulation moments obey:*

i. $\mathbb{E}^{W,\hat{W}}[W_t \hat{W}_s] = 0$, *for all s, t where $s \neq t$.*

ii. $\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}] = 0$, *for all $w_{1:t}$ and $p \geq 0$.*

Claim (ii) in Lemma 5.1 is important. It states that the proxy estimators of the treatment, like the treatments themselves, are uncorrelated with the future and current potential outcomes. In both cases, they can be influenced by past potential outcomes only.

5.2.3 LP-IV under additive causality

The next theorem provides a causal interpretation of the LP-IV method using the proxy potential outcome time series model. We focus on the case of additive causal effects.

Theorem 5.2 (LP-IV with shocks)**.** *Assume a proxy potential outcome time series and that*

A. The causal effects are additive, given in Definition 6.

B. $(W_t, \hat{W}_t)$ is a joint martingale difference with respect to $W_{1:t-1}, \hat{W}_{1:t-1}, Y_{1:t-1}$.

C. $\mathbb{E}^{W,\hat{W}}[W_t\hat{W}_s]$ *and* $\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t})\hat{W}_{t-p}]$ *exists for fixed* $w_{1:t}$.

D. $\mathbb{E}^{W,\hat{W}}[W_{i,t}\hat{W}_{j,t}] = 1_{i=j}\gamma_j$.

Then, for $p \geq 0$, $\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}] = \beta_{p,k,j,t}\gamma_j$.

Remark 5.3. *From Theorem 5.2, we have the following points.*

1. *Theorem 5.2 implies that the Wald estimand*

$$\frac{\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}]}{\mathbb{E}^{Y,W,\hat{W}}[Y_{j,t-p}(W_{1:t-p})\hat{W}_{j,t-p}]} = \frac{\beta_{p,k,j,t}}{\beta_{0,j,j,t}}. \tag{13}$$

   Stock and Watson (2018) a priori impose a "unit effect normalization," which restricts the contemporaneous effect of the j-th treatment on the j-th outcome to be $\beta_{0,j,j,t} = 1$.[15] *Then the ratio in Equation (13) identifies the causal coefficient of the additive causal effect* $\beta_{p,k,j,t}$.

2. *When we strengthen Assumption B to I-additive causal effects, then the Wald estimand becomes*

$$\frac{\mathbb{E}^{Y,W,\hat{W}}[Y_{k,t}(W_{1:t})\hat{W}_{j,t-p}]}{\mathbb{E}^{Y,W,\hat{W}}[Y_{j,t-p}(W_{1:t-p})\hat{W}_{j,t-p}]} = \frac{\beta_{p,k,j}}{\beta_{0,j,j}}.$$

   The usual Wald estimator is then

$$\bar{\beta}_{p,k,j} = \frac{\sum_{t=p+1}^{T}(Y_{k,t} - \bar{Y}_k)(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)}{\sum_{t=p+1}^{T}(Y_{j,t-p} - \bar{Y}_j)(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)},$$

   where we can rewrite this as

$$\bar{\beta}_{p,k,j} = \beta_{p,k,j}\frac{\sum_{t=p+1}^{T} W_{j,t-p}(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)}{\sum_{t=p+1}^{T}(Y_{j,t-p} - \bar{Y}_j)(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)} + \frac{\sum_{t=p+1}^{T}\eta_{k,j,t}(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)}{\sum_{t=p+1}^{T}(Y_{j,t-p} - \bar{Y}_j)(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)}.$$

   With additional regularity conditions, conventional asymptotic methods can be used to study the joint limiting dist of the three averages $\frac{1}{T-p}\sum_{t=p+1}^{T} W_{j,t-p}(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)$, $\frac{1}{T-p}\sum_{t=p+1}^{T}\eta_{k,j,t}(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)$ *and* $\frac{1}{T-p}\sum_{t=p+1}^{T}(Y_{j,t-p} - \bar{Y}_j)(\hat{W}_{j,t-p} - \bar{\hat{W}}_j)$. *The three equations are centered on* γ_j, 0 *and* $\beta_{0,j,j}\gamma_j$ *respectively. Once again, HAC/HAR inference techniques may be required. There is an additional worry that the measured proxy* $\hat{W}_{j,t-p}$ *is a weak instrument for the observed outcome* $Y_{j,t-p}$ *(Staiger and Stock, 1997; Andrews et al., 2018).*

[15] The unit-effect normalization is introduced as the shocks of interest are unobserved and so, it is scale normalization. Alternatively, the macroeconometric literature often adopts a unit standard deviation normalization, which imposes that $\mathbb{E}^{W}[W_tW_t'] = I_{n_w}$. See Stock and Watson (2016) for a detailed discussion of these normalizations.

We next connect the result in Theorem 5.2 to the existing literature on LP-IV in the macroeconometric literature. We show that the existing LP-IV assumptions are not sufficient to identify a dynamic causal effect. In the LP-IV literature, it is typical to consider a weaker assumption than Assumption B in Theorem 5.2. Instead of the treatment and proxy being a joint martingale difference sequences, it is common to assume that the proxy satisfies *lead-lag exogeneity*. As we formalize in Proposition B.8 of the Web Appendix, this alone is not sufficient to identify a dynamic causal effect in the potential outcome time series framework. In particular, an additional assumption, $\mathbb{E}^{Y,\hat{W}}[Y_t(\bar{w}_{1:t})\hat{W}_{t-p}] = 0$, for $p \geq 0$ is needed. This restricts the joint distribution of the unobserved counterfactual and the proxy. To the best of our knowledge, this additional necessary condition is not mentioned in the existing LP-IV literature.[16]

Remark 5.4. *In Proposition B.9 of the Web Appendix, we show that the IV estimator identifies a time-weighted average of the k, j-th lag-p time-t causal coefficients if the causal effects are additive but not I-additive. In this result, we still restrict the correlation between the proxy $\hat{W}_{j,t}$ and the treatment $W_{j,t}$ to be time invariant – that is, $\mathbb{E}^{W,\hat{W}}[W_{j,t}\hat{W}_{j,t}] = \gamma_j$ does not depend on t.*

6 Empirical illustration: nonparametrics and local projections

In this section, we compare the nonparametric approach discussed in Section 3 to the local projection approach discussed in Section 5 in a well-known empirical application that studies the causal effects of monetary policy. This is not to emphasize the differences in conclusions but rather show that the nonparametric approach is feasible, simple to implement yet requires the researcher to think carefully about the causal effects that can be feasibly estimated in the available data.

Here we replicate an analysis in Ramey (2016). We begin by reproducing local projection estimates of the causal effect of monetary policy on several macroeconomic outcomes using a well-known measure of monetary policy shocks – the series constructed in Romer and Romer (2004) ("Romer-Romer shocks").[17] We use this shock series as the observed treatments W_t in the nonparametric estimator of the weighted causal effect that we analyzed in Section 3.

[16] In private conversations, James H. Stock told us that Gary Chamberlain had mentioned the necessity of making an assumption about how the proxy and counterfactual vary together during a 2018 Harvard-MIT seminar presentation of Stock and Watson (2018).

[17] The Romer-Romer shocks are the residuals of a regression of the Federal funds target rate on Greenbook forecasts of economic conditions at each FOMC meetings. See Romer and Romer (2004) for details.

6.1 Local projection estimates

We begin by constructing local projection estimates of the causal effects of monetary policy. The analysis here reproduces Figure 3.2B in Ramey (2016).

To construct the local projection estimates, we estimate the regression specification

$$Y_{k,t} = \alpha_{p,k} + \beta_{p,k} W_{t-p} + X'_{t-p}\lambda_{p,k} + \epsilon_{p,k,t} \quad \text{for } p = 1, \ldots, P. \tag{14}$$

Here we have dropped from the notation j as we only have a single treatment at time $t-p$. $Y_{k,t}$ is the k-th macroeconomic outcome, W_{t-p} is the measured treatment, X_{t-p} is a vector of additional controls. The objects of interest are the coefficients $\beta_{p,k}$, which are commonly interpreted as tracing out the impulse response function in Definition 9 for $W_{t-p} = 1, W'_{t-p} = 0$. In the potential outcome time series, we think of $\beta_{p,k}$ as tracing out the causal coefficients in Definition 6 as p varies with $w = 1$ and $w' = 0$, provided the causal effects are I-additive.

The treatment W_{t-p} is the Romer-Romer shock. The outcomes of interest are the log of Industrial Production (IP), the log of the Consumer Price Index (CPI), the unemployment rate and the Federal funds rate. The additional controls X_{t-p} include two lags of the Romer-Romer shocks, the log of Industrial Production (IP), the log of the Consumer Price Index (CPI), the unemployment rate, the Federal funds rate and the log of the Commodity Price Index as well as contemporaneous values of the the log of Industrial Production (IP), the log of the Consumer Price Index (CPI), the unemployment rate and the log of the Commodity Price Index. The local projections using the Romer-Romer shocks are estimated over monthly data from January 1969 to December 2007. We estimate the specification in Equation (14) for $p = 1, \ldots, 24$ and report the results in Figure 2.

By the Frisch-Waugh Theorem, the local projection specification in Equation (14) is estimated using the residual variation in the measured treatments W_{t-p} net of the variation explained by the included controls X_{t-p}. As a result, we treat $\tilde{W}_{t-p}$ as the treatment of interest in the nonparametric estimates, where $\tilde{W}_{t-p} = W_{t-p} - Proj(W_{t-p}|X_{t-p})$ is the residuals of the measured treatment from a linear projection on the controls.

Figure 1 plots the observed distributions of the residualized shocks. Note how *little* remaining variation remains after we have residualized the measured shocks. Most of the data are concentrated between values of -0.25 and 0.25. Examining this distribution suggests the local projection estimates at $W_{t-p} = w = 1$, $W_{t-p} = w' = 0$ involves substantial, out-of-sample extrapolation. This extrapolation is possible due to its assumed linear functional form. On the other hand, if the chosen treatments and

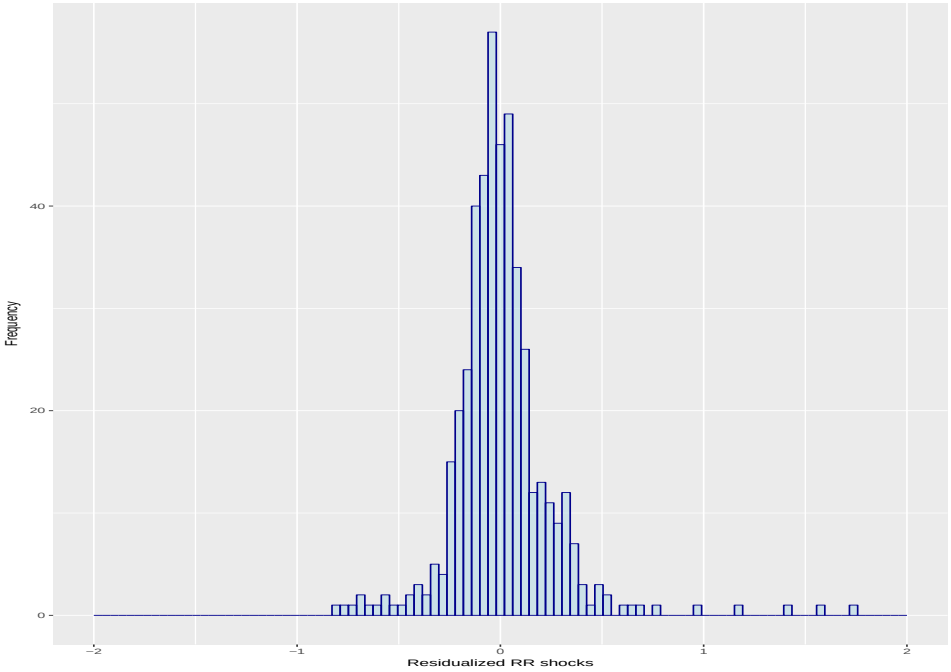


Figure 1: Romer-Romer shocks: distribution of residualized shocks: $\tilde{W}_{t-p}$. These figures report the distribution of residualized shock estimates, which are constructed from the residuals of a regression of the shock on the included controls in Equation (14).

counterfactuals lie in sparse regions of the support of $\tilde{W}_{t-p}$, the nonparametric estimates will be quite noisy. For this reason, we restrict attention to treatments and counterfactuals that lie close to zero.

6.2 Nonparametric estimates

We now turn to constructing estimates of the time-t k-th lag-p weighted causal effect $\bar{\tau}_{k,t}(w, w')(p)$ as well as the average k-th lag-p weighted causal effect $\bar{\tau}_k(w, w')(p)$. In doing so, we aim to transparently show how to translate the mechanics of the local projection estimates into the nonparametric style estimator defined in Equation (7).

The first step for the residualized Romer-Romer shocks, is to set w, w'. As mentioned earlier, we choose treatments and counterfactuals close to zero due to limited, observed variation in the data. and so, we consider the treatment and counterfactual $w = 0.1, w' = -0.1$. If the causal effects are I-additivity $\bar{\tau}_k(w, w')(p) = 0.2\beta_{p,k}$ as $w - w' = 0.2$. So, we might expect the effects we see from our nonparametrics to be only 20% those familiar from the local projection analysis.

The second step is to estimate the treatment distribution function F_{t-p} given past information. The local projection specification suggests that $\tilde{W}_t$ are difficult to predict given past data. Given this, we simply follow Hirano and Imbens (2004) and specify a parametric distribution for $\tilde{W}_{t-p}$. For

simplicity, we model $\tilde{W}_{t-p} \sim N(\gamma, \sigma^2)$ and estimate the parameters via maximum likelihood. Let $F_{t-p}(w)$ denote the distribution function of $N(\hat{\gamma}, \hat{\sigma}^2)$ evaluated at $\tilde{W}_{t-p} = w$.

The nonparametric estimator for $\bar{\tau}_{k,t}(w, w')(p)$ is then $\hat{\bar{\tau}}_k(w, w')(p)$ which is implemented as

$$\frac{1}{(T-p)} \sum_{t=p+1}^{T} \frac{(Y_{k,t} - Y_{k,t-p-1}) \left(\mathbb{1}(\tilde{W}_{t-p} \in [w-h, w+h]) - \mathbb{1}(\tilde{W}_{t-p} \in [w'-h, w'+h]) \right)}{F_{t-p}(\tilde{W}_{t-p}+h) - F_{t-p}(\tilde{W}_{t-p}-h)}. \tag{15}$$

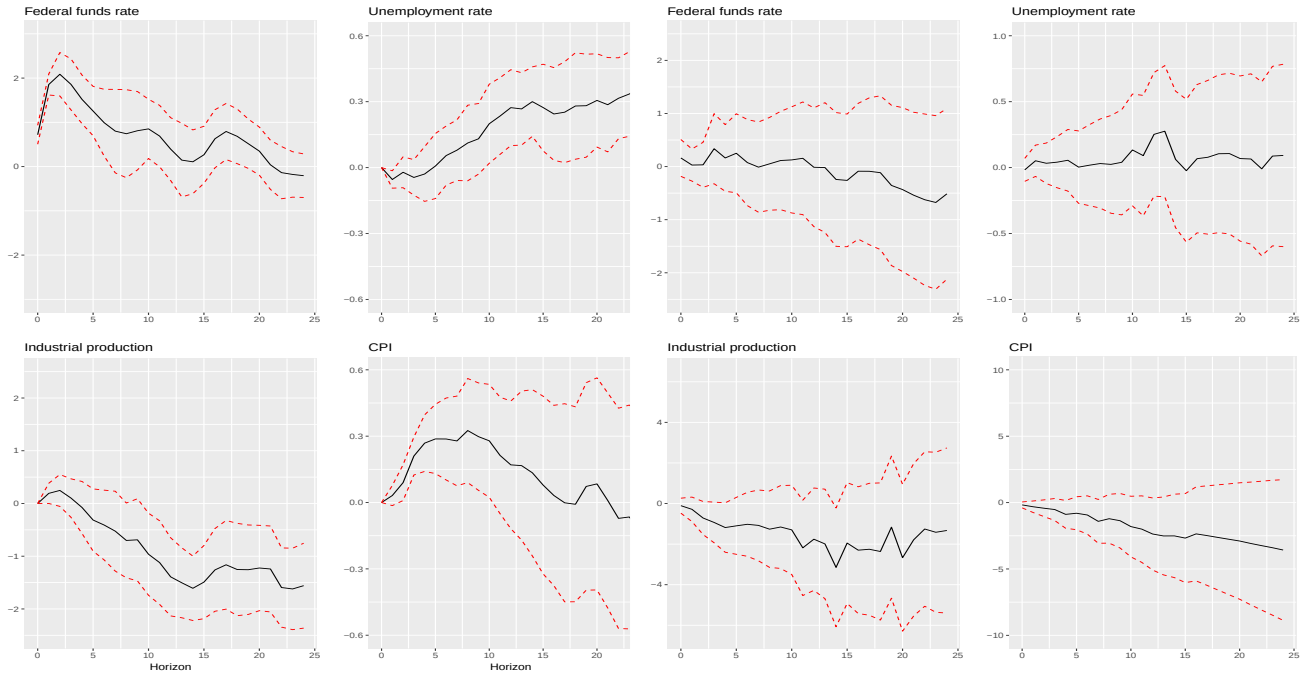
and its conservative variance estimator is $h^{-1} v_k^2(p)/(T-p)$ where

$$v_k^2(p) = \frac{1}{(T-p)} \sum_{t=p+1}^{T} \frac{(Y_{k,t} - Y_{k,t-p-1})^2 \left(\mathbb{1}(\tilde{W}_{t-p} \in [w-h, w+h]) + \mathbb{1}(\tilde{W}_{t-p} \in [w'-h, w'+h]) \right)}{f_{t-p}(\tilde{W}_{t-p})^2}. \tag{16}$$

where h is the bandwidth parameter. In practice, we allow the bandwidth to vary over k, p, so we write this as $h_k(p)$. We select it using the simple method discussed in Section 3.2.3. Note that $\hat{\bar{\tau}}_{k,t}(w, w')$ is the doubly robust estimator. This is particularly useful to further reduce the variance of the estimator and obtain informative confidence intervals.

Figure 2b plots the estimates of $\bar{\tau}_{k,j}(w, w')(p)$ for $p = 1, \ldots, 24$ of the residualized Romer-Romer shocks. It is important to emphasize that these figures plot a different object than the Figure 2a. Under the standard assumptions in the macroeconometric literature, the local projection coefficients β_p are the coefficients of the impulse response function and as we show in Theorem 5.2, they are also the causal coefficients under weaker assumptions. $\bar{\tau}_k(w, w')(p)$ is the average weighted causal effect, which is causally interpretable without any additional assumptions such as additivity or I-additivity. As mentioned earlier, if the causal effects were I-additive, we would expect the point estimates of the local projection to be about 5 times larger than the nonparametric point estimates. This does not appear to be true in this example, which suggests non-linearity in the underlying causal effects. That being said, the broad patterns appear similar – the monetary policy shock has a small positive effect on the Federal funds rate that decays over time and it slightly raises unemployment. However, the nonparametric estimates suggest that the data contain little conclusive information on the causal effect of monetary policy on industrial production and inflation as indicated by the large confidence intervals.

Taken together, this empirical application demonstrates that the nonparametric approach is both feasible and simple to implement. In particular, it weakens the typical assumptions used in empirical work and requires the researcher to be transparent about the sources of treatment variation in thee



(a) Local projection estimates

(b) Nonparametric estimates

Figure 2: Romer-Romer shocks: estimated dynamic causal effects for $p = 1, \ldots, 24$. Panel (a) plots estimated $\hat{\beta}_{p,k}$ for the LP method from Equation (14). Panel (b) plots nonparametric version: estimated $\bar{\tau}_k(w, w')(p)$. The solid black line plots the point estimates and the dotted red lines plot the point-wise 95% CI. For the LP regression, the standard errors are constructed using the Newey-West HAC estimator. Under linearity, the point estimates in Panel (a) should be 5 times larger than the estimates in Panel (b).

data. We believe that both are desirable features of the nonparametric approach.

7 A model-based approach: structural vector autoregression

As we discussed, causal inference in macroeconomics is particularly challenging because the treatment path $W_{1:t}$ is often unobserved. Another traditional approach to solving this problem is to use a model that a priori links the observed outcomes to the unobserved shocks. The chosen model thereby allows the researcher to use the observed outcomes $y_{1:T}^{obs}$ to construct estimates of dynamic causal effects without any data on the treatments. This is a very large step. Any resulting causal inferences are inevitably fragile as they rely on many a priori modelling assumptions (see Sims (2010) who strongly advocates for this extensive modelling approach).

7.1 SVAR and VAR

An influential example of this approach uses a structural vector autoregression (SVAR) to infer impulse response functions out of only observed outcomes $y_{1:T}^{obs}$. This was pioneered by Sims (1980). We now define a SVAR and the associated vector autoregression (VAR) for the potential outcome time series. This will frame the assumptions that are built into the SVAR approach and connect them to the potential outcome time series approach.

Definition 14 (SVAR and VAR)**.** *Assume a potential outcome time series and that $n_y = n_w$. Let $\Gamma_0, \ldots, \Gamma_{t-1}$ be non-stochastic $n_y \times n_y$ matrices. If for all treatment paths $w_{1:t}$ and each t,*

$$\Gamma(L)_t\{Y_t(w_{1:t}) - \mu\} = w_t, \quad \textit{where} \quad \Gamma(L)_t = \Gamma_0 - \Gamma_1 L - \ldots - \Gamma_{t-1}L^{t-1},$$

*is the associated lag polynomial and L is the lag operator, then $Y_{1:t}(\bullet)$ is a **structural vector autoregression**. Additionally, let $A_0, \ldots, A_{t-1}$ be non-stochastic $n_y \times n_y$ matrices. If for all treatment paths $w_{1:t}$ and each t,*

$$A(L)_t\{Y_t(w_{1:t}) - \mu\} = \Theta_0 w_t, \quad \textit{where} \quad A(L)_t = I - A_1 L - \ldots - A_{t-1}L^{t-1}$$

*is the associated lag polynomial, then $Y_{1:t}(\bullet)$ is a **vector autoregression**.*

The typical definition of SVAR and VAR will almost always include the assumption that $\{W_t\}$ is sequence of shocks. We will make progress without that assumption. The SVAR and VAR forms are very strong assumptions. They state that the coefficients in the polynomials are time-invariant, although here the length of the polynomial grows with t. It additionally restricts that $n_w = n_y$. That is, the number of treatments are equal to the number of outcomes and the outcomes are measured without noise. While this is a strong assumption, it is also a common, starting assumption in the existing SVAR literature.[18]

7.2 SVAR and causal equivalence

We begin with two propositions that will state several properties of the potential structural vector autoregression as well as a new property for potential outcome time series with I-additive causal effects. The results are mathematically trivial to deduce from the definitions provided, but are methodolog-

[18]While we do not discuss the case $n_w \neq n_y$, recent work in Plagborg-Møller (2019) makes progress on that case.

ically important. They will be used to establish the connections between these approaches to causal inference in time series.

Proposition 7.1. *Assume a potential outcome time series is a SVAR. Additionally assume that the matrix* Γ_0 *is invertible. Then,*

i. $Y_{1:t}$ *can be written as a VAR with* $A_s = \Gamma_0^{-1}\Gamma_s$ *for* $s = 1, 2, \ldots, t-1$ *and* $\Theta_0 = \Gamma_0^{-1}$.

ii. The time-t causal effects of $Y_{1:t}$ *follow the recursion* $A(L)_t\{Y_t(w_{1:t}) - Y_t(w'_{1:t})\} = \Theta_0(w_t - w'_t)$. *So,* $Y_{1:t}$ *has I-additive causal effects. Moreover, the time-t lag-p causal coefficients of* $Y_{1:t}$ *satisfy* $\beta_p = A_1\beta_{p-1} + ... + A_p\beta_0$ *for* $p = 1, 2, \ldots, t-1$ *with* $\beta_0 = \Theta_0$.

As SVAR time-t causal effects are I-additive, all of the earlier results we derived in Section 2 using additive causal effects apply as well. We next turn to the potential outcome time series with I-additive causal effects.

Proposition 7.2. *Assume a potential outcome time series* $Y_{1:t}$ *has I-additive causal effects and that* β_0 *is invertible. Then, the time-t causal effect follows a VAR with* $A(L)_t\{Y_t(w_{1:t}) - Y_t(w'_{1:t})\} = \beta_0(w_t - w'_t)$, *where the VAR coefficients follow the recursion* $A_p = (\beta_p - A_1\beta_{p-1} - \ldots - A_{p-1}\beta_1)\beta_0^{-1}$.

With these results, we can state the main result of this section, which proves the causal equivalence of the potential SVAR and a potential outcome time series with I-additive causal effects.

Theorem 7.1. *A structural vector autoregression with invertible* Γ_0 *is causally equivalent to a potential outcome time series with I-additive causal effects and invertible* β_0.

Proof. This follows from Proposition 7.1 and Proposition 7.2. □

Theorem 7.1 states a SVAR form with invertible Γ_0 is causally equivalent to the potential outcome time series with I-additive causal effects and an invertible matrix of contemporaneous causal coefficients β_0. Notice equivalence holds without any assumption on the treatment path.

This causal equivalence is a foundational point for macroeconometrics. As discussed in Section 4, a modern version of the Frisch (1933)-Slutzky (1937) paradigm is the shocked potential outcome time series. Hence, if we are willing to make the additional assumptions of I-additive causal effects and an invertible matrix of contemporaneous causal coefficients β_0, then we can translate the Frisch (1933)-Slutzky (1937) paradigm to a causally equivalent SVAR form with invertible Γ_0. There is *no* reason to separately introduce the structural vector moving average form. This is important for the SVMA form is often expressed as the basic primitive assumption following Sims (1980).[19]

[19] See also the introduction of Ramey (2016) for a discussion of this view.

Finally, this result is related to independent and concurrent work by Plagborg-Møller and Wolf (2019), which shows that SVARs and local projections estimate the same impulse response function in population. Provided that the causal effects are additive, Theorem 7.1 along with Theorem 5.1 show that the SVAR and LP approaches both identify the causal coefficients.

7.3 Causal equivalence and time series invertibility

Before concluding this section, we briefly discuss the connection between the causal equivalence result in Theorem 7.1 and "time series invertibility." This will clarify the context of our result in relation to the existing macroeconometric literature.

As mentioned, the existing macroeconometrics literature usually begins with the SVMA form in Definition 12 and the SVAR form provides a strategy for estimating the coefficients of the SVMA. Given that the SVMA is their primitive model, the first step in the existing literature is to provide sufficient assumptions that allow the SVMA to be transformed into an SVAR. The critical assumption in the literature is *time series invertibility* (Kilian and Lutkepohl, 2017; Stock and Watson, 2018).

Definition 15 (Time series invertibility)**.** *The SVMA in Definition 12 is* ***invertible*** *if* $n_y = n_w$ *and the lag polynomial* $\Theta(L)_t = \Theta_0 - \Theta_1 L - \ldots - \Theta_t L^t$ *is invertible.*

If time series invertibility holds, then the SVMA in Definition 12 can be transformed into the SVAR in Definition 14 by inverting the lag polynomial and vice versa. Stock and Watson (2018) show that, in the context of linear models such as the SVMA, time series invertibility is equivalent to $\eta_t = \Theta_0 w_t$ and Θ_0^{-1} existing, where η_t is the reduced form errors from the VAR.

While it may appear that time series invertibility is a stronger assumption than the conditions in Theorem 7.1, they are equivalent and any difference is due to our focus on the finite case with $T < \infty$. To see this, recall from Section 4 that the SVMA form is causally equivalent to any I-additive potential outcome time series by setting $\Theta_s = \beta_s$ for $s = 1, \ldots, t$. With this in mind, Theorem 7.1 can be re-stated as: A SVAR form with invertible Γ_0 is causally equivalent to the SVMA form with an invertible Θ_0. The details of the SVAR and VAR forms in Definition 14 impose the restriction that the reduced form errors from the VAR are a linear combination the treatments. As a result, the assumption that $\Theta_0 = \beta_0$ is invertible is then equivalent to invertibility in Definition 15.

So, the typical invertibility assumption in the literature implies the causal equivalence between the SVAR and the potential outcome time series with I-additive causal effects.

8 Conclusion

In this paper, we established a nonparametric foundation to formalize dynamic causal effects in time series data. We did so by extending the potential outcome time series to suit the observational time series settings that are common in macroeconomics. We used this framework to develop nonparametric estimators of dynamic causal effects as well as formalize the causal content of several common estimation strategies in macroeconometrics.

In Section 2, we defined a dynamic causal effect as the comparison of potential outcomes at a fixed point in time along different treatment paths. These dynamic causal effects served as the basis for a series of causal estimands such as the weighted causal effect, the causal response function and the impulse response function. These estimands differed from one another as they average over different objects. For example, the weighted causal effect held the potential outcomes fixed, while the causal response function treated both the treatments and potential outcomes as random. In defining these estimands, we distinguished between superpopulation causal effects and sample causal effects, a difference that is well understood and frequently discussed in the microeconometris literature.

Our key identifying assumption in this paper is non-anticipation (Assumption 2). That is, the current treatments are independent of future potential outcomes conditional on past outcomes and treatments. With this assumption in hand, we turned to the nonparametric identification of the dynamic causal estimands. In Section 3, we introduced a nonparametric estimator for dynamic causal effects. This estimator is unbiased and we developed tools for exact and asymptotic inference under the randomization distribution. This work illustrated how the potential outcome time series model can be used to develop new estimation methods for dynamic causal effects. At a more fundamental level, our nonparametric estimator highlights that non-anticipation is sufficient to identify dynamic causal estimands in a potential outcome time series.

We placed the potential outcome time series framework into the context of existing work in macroeconometrics. In particular, we considered treatment paths that are sequences of shocks in Section 4 and provided a definition that formalizes the common heuristic that shocks are unpredictable. We showed that assuming that the treatments are shocks together with an assumption that the causal effects are additive dramatically simplifies causal analysis on time series data as under them, several dynamic causal estimands are equivalent. However, neither of these assumptions are necessary for identification. Clarifying the role of these assumptions is an important contribution of this paper.

Finally, we argued that the shocked potential outcome time series can be thought of as a nonparametric causal formalization of the Frisch (1933)-Slutzky (1937) paradigm and therefore be used

to deepen our understanding of existing estimation methods. For example, in Section 5, we provided a formal causal interpretation of LP and LP-IV. We showed that if the causal effects are additive and some additional conditions hold, LP identifies the causal coefficients that trace out the lag-p causal effects and if the causal effects are also I-additive, this can be directly implemented via ordinary least squares. We also showed that the existing LP-IV conditions in the literature are not sufficient to identify a causal effects in the potential outcome time series model and provided an alternative set of sufficient conditions. We discussed SVARs in Section 7 and showed that an SVAR is causally equivalent to an I-additive potential outcome time series under the usual time series invertibility condition.

Taken together, we believe that nonparametric potential outcome time series provides a unified foundation for the analysis of dynamic causal effects in macroeconomics.

References

Abadie, A., S. C. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report.

Abadie, A., S. C. Athey, G. W. Imbens, and J. Wooldridge (2018). Sampling-based vs. design-based uncertainty in regression analysis. Technical report.

Abadie, A. and M. D. Cattaneo (2018). Econometric methods for program evaluation. *Annual Review of Economics 10*, 465–503.

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, pp. 5145–5303. Amsterdam, The Netherlands: North Holland.

Abbring, J. H. and G. J. van den Berg (2003). The nonparametric identification of treatment effects in duration models. *Econometrica 71*, 1491–1517.

Andrews, I., J. H. Stock, and L. Sun (2018). Weak instruments in IV regression: theory and practice. Technical report.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*, 444–455.

Angrist, J. D., O. Jorda, and G. M. Kuersteiner (2018). Semiparametric estimates of monetary policy effects: string theory revisited. *Journal of Business and Economic Statistics 36*, 381–387.

Angrist, J. D. and G. M. Kuersteiner (2011). Causal effects of monetary shocks: Semiparametric conditional independence tests with a multinomial propensity score. *Review of Economics and Statistics 93*, 725–747.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Angrist, J. D. and J.-S. Pischke (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives 24*, 3–30.

Aronow, P. M. and C. D. Samii (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science 60*, 250–267.

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

Bang, H. and J. M. Robins (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics 61*, 962–973.

Bernanke, B. S. and K. Kuttner (2005). What explains the stock market's reaction to Federal Reserve policy? *Journal of Finance 60*, 1221–1257.

Blackwell, M. and A. Glynn (2018). How to make causal inferences with time-series and cross-sectional data. *American Political Science Review 112*, 1067–1082.

Bojinov, I. and N. Shephard (2019). Time series experiments and causal estimands: exact randomization tests and trading. *Journal of the American Statistical Association*. Forthcoming.

Bondersen, K. H., F. Gallusser, J. Koehler, N. Remy, and S. L. Scott (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics 9*, 247–274.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018). Diagnostic expectations and credit cycles. *Journal of Finance 73*, 199–227.

Boruvka, A., D. Almirall, K. Witkiwitz, and S. A. Murphy (2018). Assessing time-varying causal effect moderation in mobile health. *Journal of the American Statistical Association 113*, 1112–1121.

Box, G. E. P. and G. C. Tiao (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association 70*, 70–79.

Brodersen, K., F. Gallusser, J. Koehler, N. Remy, and S. Scott (2015). Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics 9*, 247–274.

Busso, M., J. DiNardo, and J. McCrary (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics 96*, 885–897.

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

Chamberlain, G. (1982). The general equivalence of Granger and Sims causality. *Econometrica 50*, 1305–1324.

Christiano, L. J., M. S. Eichenbaum, and C. L. Evans (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1A, pp. 65–148. Amsterdam, The Netherlands: North Holland.

Christiano, L. J., M. S. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy 113*, 1–45.

Cochrane, J. H. and M. Piazessi (2002). The Fed and interest rates - a high-frequency identification. *American Economic Review 92*, 90–95.

Cox, D. R. (1958). *Planning of Experiments*. Oxford, United Kingdom: Wiley.

Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. *Journal of the Royal Statistical Society, Series B 23*, 414–422.

D'Amour, A., P. Ding, A. Feller, L. Lei, and J. Sekhon (2018). Overlap in observational studies with high-dimensional covariates. *arXiv:1711.02582 preprint*.

Dawid, A. P. (2000). Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association 95*, 407–448.

Durbin, J. (1954). Errors in variables. *Review of the Institute of International Statistics 22*, 23–54.

Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica 51*, 277–304.

Engle, R. F., T. Ito, and W. L. Lin (1990). Meteor showers or heat waves? Heteroskedastic intra-day volatility in the foreign exchange market. *Econometrica 58*, 525–542.

Faust, J., J. H. Rogers, E. Swanson, and J. H. Wright (2003). Identifying the effects of monetary policy shocks on exchange rates using high frequency data. *Journal of the European Economic Association 1*, 1031–1057.

Fieldhouse, A., K. Mertens, and M. Ravn (2018). The macroeconomic effects of government asset purchases: evidence from postwar U.S. housing credit policy. *Quarterly Journal of Economics 133*, 1503–1560.

Fisher, R. A. (1925). *Statistical Methods for Research Workers* (1 ed.). London, United Kingdom: Oliver and Boyd.

Fisher, R. A. (1935). *Design of Experiments* (1 ed.). London, United Kingdom: Oliver and Boyd.

Frangakis, C. E. and D. B. Rubin (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika 86*, 365–379.

Frisch, R. (1933). *Propagation Problems and Impulse Problems in Dynamic Economics*. London, United Kingdom: Allen and Unwin.

Fuster, A., D. I. Laibson, and B. Mendel (2010). Natural expectations and macroeconomic fluctuations. *Journal of Economic Perspectives 24*, 67–84.

Gallant, A. R., P. E. Rossi, and G. Tauchen (1993). Nonlinear dynamic structures. *Econometrica 61*, 871–907.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*, 424–438.

Gurkaynak, R. S., B. Sack, and E. Swanson (2005). The sensitivity of long-term interest rates to economic news: evidence and implications for macroeconomic models. *American Economic Review 95*, 425–436.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications*. San Diego, California, USA: Academic Press.

Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics 113*, 363–398.

Hansen, B. E. (2018). *Econometrics*. web version, downloaded 31 October 2018.

Harvey, A. C. (1996). Intervention analysis with control groups. *International Statistical Review 64*, 313–328.

Harvey, A. C. and J. Durbin (1986). The effects of seat belt legislation on British road casualties: A case study in structural time series modelling. *Journal of the Royal Statistical Society, Series A 149*, 187–227.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016). Dynamic treatment effects. *Journal of Econometrics 191*, 276–292.

Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*, 341–396.

Hendry, D. F. (2017). Granger causality. *European Journal of Pure and Applied Mathematics 10*, 12–29.

Hennessy, J., T. Dasgupta, L. Miratrix, C. Pattanayak, and P. Sarkar (2016). A conditional randomization test to account for covariate imbalance in randomized experiments. *Journal of Causal Inference 4*, 61–80.

Herbst, E. and F. Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton, New Jersey, USA: Princeton University Press.

Hernan, M. A. and J. M. Robins (2019). *Causal Inference*. Boca Raton, Florida, USA: Chapman & Hall. Forthcoming.

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pp. 73–84. Hoboken, New Jersey, USA: John Wiley.

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

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*, 663–685.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*, 467–475.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference for statistics, social and biomedical sciences: an introduction*. Cambridge, United Kingdom: Cambridge University Press.

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

Jordá, O. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*, 161–182.

Jordá, O., M. Schularick, and A. M. Taylor (2015). Betting the house. *Journal of International Economics 96*, S2–S18.

Justiniano, A. and G. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review 98*, 604–641.

Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of the American Statistical Association 50*, 946–967.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*, 2021–2042.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2002). Simple robust testing of regression hypotheses. *Econometrica 68*, 695–714.

Kilian, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature 46*, 871–909.

Kilian, L. (2011). Structural vector autoregressions. Technical report, C.E.P.R. Discussion Papers.

Kilian, L. and H. Lutkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge, United Kingdom: Cambridge University Press.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*, 119–147.

Kuersteiner, G. (2010). Granger-Sims causality. In S. N. Durlauf and L. Blume (Eds.), *Macroeconomics and Time Series Analysis*, pp. 119–134. London, United Kingdom: Palgrave Macmillian.

Kuersteiner, G., D. Phillips, and M. Villamizar-Villegas (2018). Effective sterilized foreign exchange intervention? Evidence from a rule-based policy. *Journal of International Economics 118*, 118–138.

Kuttner, K. (2001). Monetary policy surprises and interest rates: evidence from the Fed funds futures market. *Journal of Monetary Economics 47*, 523–544.

Lazarus, E., D. J. Lewis, J. H. Stock, and M. W. Watson (2018). HAR inference: recommendations for practice (with discussion). *Journal of Business and Economic Statistics 36*, 541–559.

Lechner, M. (2011). The relation of different concepts of causality used in time series and microeconomics. *Econometric Reviews 30*, 109–127.

Lu, X., L. Su, and H. White (2017). Granger causality and structural causality in cross-section and panel data. *Econometric Theory 33*, 263–291.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory 4*, 103–124.

Marcellino, M., J. H. Stock, and M. W. Watson (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics 135*, 499–526.

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics 22*, 1142–1160.

Mertens, K. and J. L. M. Olea (2018). Marginal tax rates and income: new time series evidence. *The Quarterly Journal of Economics 133*, 1803–1884.

Mertens, K. and M. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review 103*(4), 1212–1247.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*, 1589–1619.

Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statisticsl Society B 65*, 331–366.

Murphy, S. A., M. J. van der Laan, J. M. Robins, and C. P. P. R. Group (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association 96*, 1410–1423.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica 29*, 315–335.

Nakamura, E. and J. Steinsson (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics 133*, 1283–1330.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science 5*, 465–472. Originally published 1923, republished in 1990, translated by Dorota M. Dabrowska and Terence P. Speed.

Plagborg-Møller, M. (2019). Bayesian inference for structural impulse response functions. *Quantitative Economics*. Forthcoming.

Plagborg-Møller, M. and C. K. Wolf (2018). Instrumental variable identification of dynamic variance decompositions. Unpublished paper: Department of Economics, Princeton University.

Plagborg-Møller, M. and C. K. Wolf (2019). Local projections and VARs estimate the same impulse responses. Unpublished paper: Department of Economics, Princeton University.

Priestley, M. B. (1988). *Nonlinear and Non-stationary Time Series Analysis*. London, United Kingdom: Academic Press.

Ramey, V. A. (2016). Macroeconomics shocks and their propagation. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2A, pp. 71–162. Amsterdam, The Netherlands: North Holland.

Ramey, V. A. and S. Zubairy (2018). Government spending multipliers in good times and in bad: evidence from US historical data. *Journal of Political Economy 126*, 850–901.

Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *Journal of the American Statistical Association 85*, 132–138.

Ricciardi, F., A. Mattei, and F. Mealli (2016). Bayesian inference for sequential treatments under latent sequential ignorability. Unpublished paper: Department of Statistics, Universita degli Studi di Firenze, Italy.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods: application to control of the healthy worker survivor effect. *Mathematical Modelling 7*, 1393–1512.

Robins, J. M. (1994). Correcting for non-compliance in randomization trials using structural nested mean models. *Communications in Statistics — Theory and Methods 23*, 2379–2412.

Robins, J. M., S. Greenland, and F.-C. Hu (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association 94*, 687–700.

Robins, J. M., M. A. Hernan, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*, 550–560.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*, 846–866.

Romer, C. D. and D. H. Romer (1989). Does monetary policy matter? A new test in the spirit of Friedman and Schwartz. *NBER Macroeconomics Annual*, 122–170.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: derivation and implications. *American Economic Review 94*, 1055–1084.

Romer, C. D. and D. H. Romer (2010). The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review 100*, 763–801.

Rosenbaum, P. R. (2002). Covariance adjustment in randomized experiments and observational studies. *Statistical Science 17*, 286–304.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*, 135–146.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology 66*, 688–701.

Rudebusch, G. D. (1998). Do measures of monetary policy in a VAR make sense? *International Economic Review 39*, 907–931.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica 26*, 393–415.

Sargent, T. J. (1981). Interpreting economic time series. *Journal of Political Economy 89*, 213–248.

Savage, I. R. (1954). *The Foundations of Statistics*. New York, USA: Dover.

Silverman, B. W. (1986). *Density Estimation for Statistical and Data Analysis*. London, United Kingdom: Chapman & Hall.

Sims, C. A. (1972). Money, income and causality. *American Economic Review 62*, 540–552.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*, 1–48.

Sims, C. A. (2010). But economics is not an experimental science. *Journal of Economic Perspectives 24*, 59–68.

Slutzky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica 5*, 105–146.

Smets, F. R. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association 1*, 1123–1175.

Smets, F. R. and R. Wouters (2007). Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review 97*, 586–606.

Staiger, D. O. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H. and M. W. Watson (2003). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*, 159–230.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2A, pp. 415–525.

Stock, J. H. and M. W. Watson (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *Economic Journal 128*, 917–948.

Theil, H. (1971). *Principles of Econometrics*. New York, USA: John Wiley & Sons.

Toulis, P. and D. C. Parkes (2016). Long-term causal effects via behavioral game theory. 30th Conference on Neural Information Processing Systems (NIPS'16).

White, H. and X. Lu (2010). Granger causality and dynamic structural systems. *Journal of Financial Econometrics 8*, 193–243.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenbeck (Ed.), *Modern Mathematics*, pp. 165–190. New York, USA: McGraw-Hill.

Wu, J. and P. Ding (2018). Randomization tests for weak null hypotheses. Department of Statistics, University of California at Berkeley.

Yang, S., G. W. Imbens, Z. Cui, D. E. Faries, and Z. Kadziola (2016). Propensity score matching and subclassification in observational studies with multilevel treatments. *Biometrics 72*, 1055–1065.

A Mathematical appendix

This appendix provides the proofs for the main results in the body of the paper.

A.1 Results in Section 3: nonparametric estimation

Proof of Theorem 3.1

To focus on ideas, suppress in the notation references to t, p and conditioning variables w_{t-p-1}^{obs} and $\mathcal{F}_{T,t-p-1}$ as these are held constant. Write $W = W_{j,t-p}$ and $U = W_{1:j-1,t-p}, W_{j+1:n_w,t-p}, W_{t-p+1:t}$. Now sample V from $p(w,u)$ then the Horvitz and Thompson (1952) type estimator we study takes on the form

$$\hat{\tau} = \sum_u \{c_{w,u} 1_{V=(w,u)} - c_{w',u} 1_{V=(w',u)}\}$$

where $c_{w,u} = y(w,u)/p(w)$ is non-stochastic (that is V is the only random object here). The estimand is the weighted causal effect

$$\bar{\tau} = \sum_u \{y(w,u)p(u|w) - y(w',u)p(u|w')\}$$

Now, as $p(u|w) = p(w,u)/p(w)$, we immediately see that

$$\mathbb{E}^V(\hat{\tau}) = \sum_u \{c_{w,u}p(w,u) - c_{w',u}p(w',u)\} = \sum_u \{y(w,u)p(u|w) - y(w',u)p(u|w')\}$$

So the estimator is conditionally unbiased and this shows the first result. It has error

$$R = \hat{\tau} - E(\hat{\tau}) = \sum_u \left\{ c_{w,u}(1_{V=(w,u)} - p(w,u)) - c_{w',u}(1_{V=(w',u)} - p(w',u)) \right\}.$$

This has zero mean and variance $Var(R)$, which is the double sum $\sum_u \sum_{u'}$ of

$$\begin{aligned}
&c_{w,u}c_{w,u'}E[(1_{V=(w,u)} - p(w,u))(1_{V=(w,u')} - p(w,u'))] \\
&-c_{w,u}c_{w',u'}E[(1_{V=(w,u)} - p(w,u))(1_{V=(w',u')} - p(w',u'))] \\
&-c_{w',u}c_{w,u'}E[(1_{V=(w',u)} - p(w',u))(1_{V=(w,u')} - p(w,u'))] \\
&+c_{w',u}c_{w',u'}E[(1_{V=(w',u)} - p(w',u))(1_{V=(w',u')} - p(w',u'))].
\end{aligned}$$

Then note that

$$\begin{aligned}
&E[(1_{V=(w,u)} - p(w,u))(1_{V=(w,u')} - p(w,u'))] \\
&= 1_{u=u'}p(w,u) - p(w,u)p(w,u') - p(w,u)p(w,u') + p(w,u)p(w,u') \\
&= 1_{u=u'}p(w,u) - p(w,u)p(w,u').
\end{aligned}$$

So, we conclude

$$E[(1_{V=(w',u)} - p(w',u))(1_{V=(w',u')} - p(w',u'))] = 1_{u=u'}p(w',u) - p(w',u)p(w',u').$$

Likewise, if $w \neq w'$, then

$$\begin{aligned}
&E[(1_{V=(w,u)} - p(w,u))(1_{V=(w',u')} - p(w',u'))] \\
&= -p(w,u)p(w',u') - p(w,u)p(w',u') + p(w,u)p(w',u') = -p(w,u)p(w',u')
\end{aligned}$$

and similarly,

$$E[(1_{V=(w',u)} - p(w',u))(1_{V=(w,u')} - p(w,u'))] = -p(w,u')p(w',u).$$

Thus,

$$\begin{aligned}
&Var(R) = \textstyle\sum_u \left(c_{w,u}^2 p(w,u) + c_{w',u}^2 p(w',u) \right) \\
&- \textstyle\sum_u \sum_{u'} \Big(c_{w,u}c_{w',u'}p(w,u)p(w',u') + c_{w',u}c_{w,u'}p(w',u)p(w,u') \\
&+ c_{w',u}c_{w',u'}p(w',u)p(w',u') + c_{w,u}c_{w,u'}p(w,u)p(w,u') \Big).
\end{aligned}$$

Now,

$$c_{w,u}^2 p(w,u) = \frac{p(u|w)y(w,u)^2}{p(w)}, \quad c_{w',u}^2 p(w',u) = \frac{p(u|w')y(w',u)^2}{p(w')}, \quad c_{w,u}p(w,u) = p(u|w)y(w,u).$$

So, putting this all together,

$$\begin{aligned}
&Var(R) = \textstyle\sum_u \left(\frac{p(u|w)y(w,u)^2}{p(w)} + \frac{p(u|w')y(w',u)^2}{p(w')} \right) \\
&- \textstyle\sum_u \sum_{u'} \Big(p(u|w)y(w,u)p(u'|w')y(w',u') - p(u|w')y(w',u)p(u'|w)y(w,u') \\
&- p(u|w')y(w',u)p(u'|w')y(w',u') + p(u|w)y(w,u)p(u'|w)y(w,u') \Big).
\end{aligned}$$

$$Var(R) = \sum_u \left(p(u|w) \frac{y(w,u)^2}{p(w)} + p(u|w') \frac{y(w',u)^2}{p(w')} \right)$$
$$- \left(\sum_u p(u|w) y(w,u) - p(u|w') y(w',u) \right) \left(\sum_{u'} p(u'|w) y(w,u') - p(u'|w') y(w',u') \right),$$

and the result follows from the definition of $\hat{\tau}$.

The CLT result is now relatively simple. It follows from an application of the triangular martingale array central limit theory given in Theorem 3.2 of Hall and Heyde (1980). The Lindeberg condition must hold due to compactness of the set of possible treatments and the conditioning on the potential outcome paths.

A.2 Result in Section 4: Frisch-Slutzky paradigm

Proof of Theorem 4.1

First we establish the following general result.

Proposition A.1. *Assume a potential outcome time series $Y_{1:t}(\bullet)$ with additive causal effects. Then,*

i. Weighted causal effect is $\bar{\tau}_{k,j,t}(w,w')(p) = \beta_{p,k,j,t}(w - w') + \sum_{s=1}^{p} \sum_{i=1}^{n_w} \beta_{s,k,i,t} \bar{V}_{s,i,j,t-p}(w,w')$, with $\bar{V}_{s,i,j,t}(w,w') = \mathbb{E}^{W \,|\, Y(\bullet)}[W_{i,t+s} \,|\, w^{obs}_{1:t-1}, W_{j,t} = w] - \mathbb{E}^{W \,|\, Y(\bullet)}[W_{i,t+s} \,|\, w^{obs}_{1:t-1}, W_{j,t} = w']$.

ii. Causal response function is $CRF_{k,j,t}(w,w')(p) = \beta_{p,k,j,t}(w-w') + \sum_{s=1}^{p} \sum_{i=1}^{n_w} \beta_{s,k,i,t} \dot{V}_{s,i,j,t-p}(w,w')$, with $\dot{V}_{s,i,j,t}(w,w') = \mathbb{E}[W_{i,t+s} \,|\, y^{obs}_{1:t-1}, w^{obs}_{1:t-1}, W_{j,t} = w] - \mathbb{E}[W_{i,t+s} \,|\, y^{obs}_{1:t-1}, w^{obs}_{1:t-1}, W_{j,t} = w']$.

Proof. Recall the form of $\tau_{k,j,t}(w,w')(p)$ from Proposition 2.1. Applying the definitions of the weighted and dynamic causal effects produces the results in claims (i)-(ii) by conditional expectations. □

Now apply the martingale property of the treatments to Proposition A.1. The MD property implies that all the $\dot{V}_{s,i,j,t}(w,w')$ and $\bar{V}_{s,i,j,t}(w,w')$ terms are zero. This delivers the result.

A.3 Results in Section 5: Local projections

Proof of Theorem 5.1

For a potential outcome time series, imagine a fixed known counterfactual $\bar{w}_{1:t} = 0$. Then,

$$Y_t(W_{1:t}) = Y_t(\bar{w}_{1:t}) + \tau_t(W_{1:t}, \bar{w}_{1:t}).$$

By the Assumption A of additive causal effects and by the Assumption B of the existence of the mean of the counterfactual outcome, we can decompose

$$Y_{k,t}(W_{1:t}) = \mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t})] + \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}, \tag{17}$$

where

$$\eta_{k,j,t} = Y_{k,t}(\bar{w}_{1:t}) - \mathbb{E}^Y[Y_{k,t}(\bar{w}_{1:t})] + \textstyle\sum_{i=1;i\neq j}^{n_w} \beta_{p,k,i} W_{i,t-p} + \sum_{s=0;s\neq p}^{t-1} \sum_{i=1}^{n_w} \beta_{s,k,i,t} W_{i,t-s}. \tag{18}$$

The shock property means the treatments are martingale differences and so, if it exists, $\mathbb{E}^{W,Y}(\eta_{k,j,t}) = 0$ for all k, j, t. Now

$$\begin{aligned}
&\mathbb{E}^{Y,W}(Y_t(\bar{w}_{1:t}) W_{t-p}) = \mathbb{E}^Y\{Y_t(\bar{w}_{1:t}) \mathbb{E}^{W|Y(\bullet)}(W_{t-p}|Y_t(\bar{w}_{1:t}), Y_{t-p-1}(\bar{w}_{1:t}))\}, \quad \text{by iterated expectation,} \\
&= \mathbb{E}^Y\{Y_t(\bar{w}_{1:t}) \mathbb{E}^{W|Y(\bullet)}(W_{t-p}|Y_{t-p-1}(\bar{w}_{1:t}))\}, \quad \text{by non-anticipating treatments,} \\
&= 0, \quad \text{by MD assumption on treatments.}
\end{aligned} \tag{19}$$

The properties of a shock in Definition 10 then imply that $Cov^{W,Y}(W_{j,t-p}, \eta_{k,j,t}) = 0$, if this covariance exists. The existence of the moments is guaranteed by Assumption C.

Proof of Lemma 5.1

Claim (i) is standard from the martingale difference assumption. Claim (ii)

$$\begin{aligned}
&\mathbb{E}^{Y,\hat{W}}[Y_t(w_{1:t}) \hat{W}_{t-p}] \\
&\overset{(1)}{=} \mathbb{E}^Y\{Y_t(w_{1:t}) \mathbb{E}^{\hat{W}|Y}[\hat{W}_{t-p}|Y_{1:t}(w_{1:t})]\} \\
&\overset{(1)}{=} \mathbb{E}^Y\{Y_t(w_{1:t}) \mathbb{E}^{W_{1:t-p},\hat{W}_{1:t-p-1}|Y} \mathbb{E}^{\hat{W}|Y,W}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p}, Y_{1:t}(w_{1:t})]\} \\
&\overset{(2)}{=} \mathbb{E}^Y\{Y_t(w_{1:t}) \mathbb{E}^{W_{1:t-p},\hat{W}_{1:t-p-1}|Y} \mathbb{E}^{\hat{W}|Y,W}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p}, Y_{1:t-p-1}(w_{1:t-p-1})]\} \\
&\overset{(3)}{=} \mathbb{E}^Y\{Y_t(w_{1:t}) \mathbb{E}^{W_{1:t-p-1},\hat{W}_{1:t-p-1}|Y} \mathbb{E}^{\hat{W}|Y}[\hat{W}_{t-p}|\hat{W}_{1:t-p-1}, W_{1:t-p-1}, Y_{1:t-p-1}(w_{1:t-p-1})]\} \\
&\overset{(4)}{=} 0,
\end{aligned} \tag{20}$$

where (1) is by iterated expectations, (2) is by non-anticipation of $\hat{W}_t$, (3) integrates out W_{t-p} and (4) is by the martingale differences assumption.

Proof of Theorem 5.2

Using the definition of additive causality in Definition 6, for any non-stochastic counter-factual $\bar{w}_{1:t}$,

$$\begin{aligned} &Y_{k,t}(W_{1:t}) = \beta_{p,k,j,t} W_{j,t-p} + \eta_{k,j,t}, \quad \text{where} \\ &\eta_{k,j,t} = Y_{k,t}(\bar{w}_{1:t}) + \textstyle\sum_{i=1;i\neq j}^{n_w} \beta_{p,k,i,t} W_{i,t-p} + \sum_{s=0;s\neq p}^{t-1} \sum_{i=1}^{n_w} \beta_{s,k,i,t} W_{i,t-s}. \end{aligned} \tag{21}$$

Then, the result follows by checking off all the terms using Assumption D and using the results from Lemma 5.1, which can be used under the assumptions of this Theorem. □

A.4 Results in Section 7: Structural vector autoregression

Proof of Proposition 7.1

Claim (i) is immediate from Definition 14. We now prove claim (ii). From the definition of the VAR,

$$A(L)_t\{y_t(w_{1:t}) - \mu\} - A(L)_t\{y_t(w'_{1:t}) - \mu\} = \Theta_0(w_t - w'_t).$$

But the left hand side simplifies, as μ cancels. We are left with

$$A(L)_t\{y_t(w_{1:t}) - y_t(w'_{1:t})\} = \Theta_0(w_t - w'_t)$$

as claimed. To be explicit about the solution, write $\tau_t = y_t(w_{1:t}) - y_t(w'_{1:t})$

$$\begin{aligned} \tau_1 &= \Theta_0(w_1 - w'_1) \\ \tau_2 &= A_1\{y_1(w_1) - y_1(w'_1)\} + \Theta_0(w_2 - w'_2) \\ &= \Theta_0(w_2 - w'_2) + A_1\Theta_0(w_1 - w'_1) \\ \tau_3 &= A_1\{y_2(w_{1:2}) - y_2(w'_{1:2})\} + A_2\{y_1(w_1) - y_1(w'_1)\} + \Theta_0(w_3 - w'_3) \\ &= \Theta_0(w_3 - w'_3) + A_1\{\Theta_0(w_2 - w'_2) + A_1\Theta_0(w_1 - w'_1)\} + A_2\{\Theta_0(w_1 - w'_1)\} \\ &= \Theta_0(w_3 - w'_3) + A_1\beta_0(w_2 - w'_2) + (A_1\beta_1 + A_2\beta_0)(w_1 - w'_1), \end{aligned}$$

and so on. Reading off the coefficients in front of $w_t - w'_t$ gives us the result in claim (ii). □

Proof of Proposition 7.2

This follows from the recursion in claim (ii) of Proposition 7.1. We now write

$$A_p\beta_0 - \beta_p = A_1\beta_{p-1} - \ldots - A_{p-1}\beta_1.$$

With β_0 invertible, we arrive at the claimed result. □