REAL-TIME CAUSAL INFERENCE

BY KWEKU OPOKU-AGYEMANG

Center for Effective Global Action
Department of Agricultural and Resource Economics
University of California, Berkeley
Berkeley, California, U.S.A.
and Cornell Tech, New York, U.S.A.
kweku@berkeley.edu

Abstract: The paper highlights causal inference based on econometric measurement in real-time data environments. Each state has a probability of being realized in real-time. We define state selection bias as arising when real-time environments are ignored. We model indicator variables as measurements that exist partly in all particular theoretically possible states, but show only one configuration on observation. Under real-time randomization within data streams, econometric treatment effects are estimable using controlled and natural experiments motivated by real-time regression analyses. A bias occurs as a result of ignoring concept drift when classical regression statistics are naïvely applied to real-time experimental data. We present a simple algorithm for difference-in-difference estimation for real-time program evaluations. Finally, a new Problem of Causal Inference is introduced for real-time data environments.

Keywords: Causal inference; Real-Time Data; Randomized Experiments; Natural Experiments.

JEL: C10, C18.

Acknowledgments: I thank several people at the Center for Effective Global Action, the Department of Agricultural and Resource Economics, the Berkeley Institute for Data Science, the Berkeley Institute for Transparency in Social Science, and the Mechanical Engineering Department, all of the University of California, Berkeley as well as the Working Group in African Political Economy for various informal discussions. The paper did not use any grant funding. The usual disclaimer applies.

1. Introduction

This paper discusses approaches of measurement and causal inference within the context of real-time data, which is information delivered immediately after collection. Such data is increasingly ubiquitous in a rapidly growing segment of the knowledge economy that collects vast amounts of consumer and firm information. Examples include networked sensors in smart phone devices, social media platforms, personalized mobile health technologies and ride-sharing initiatives, all of which are based on the premise of nearly constantly generating data. To understand how to analyze such information, the paper is motivated by important results in data streams analyses in computer science and engineering (Zhu and Shaha, 2002, Babcock et al 2002, Nadungodage et al 2011), but that have not yet been harnessed within the econometrics of causal inference literature. In such settings, there is little to no delay in the timeliness of the data as well as the statistical and computational problems of data measurement and analysis and experimenters must collect data and create model parameters instantaneously. Real-time variation has not yet been discussed in the econometric program evaluation literature, although real-time information processing explains bounded-rationality (e.g. Van Zandt (1999)) and may lead to different macroeconomic policy rules from rules generated from ex-post revised data (Orphanides, 2001).

The paper builds on a growing number of methods are arising to analyze high-dimensional statistical data environments (e.g. Belloni and Chernozhukov (2009); Belloni, Chen, Chernozhukov, and Hansen (2012); Belloni, Chernozhukov, Hansen, and Fernandez-Val (2013a); Belloni, Chernozhukov, and Hansen (2013b); Liran and Einav (2014); King (2014) and Varian (2014)). Real-time data may also be important for measurement processes in quantum computing (e.g. Dragoman and Dragoman 2004), as well as technology-driven data collection (Couper, 2005) or experimental settings where study subjects have significant agency (Chassang et al, 2010) among others[1]. An important but understudied phenomenon in the empirical literature is that being surveyed, in and of itself, can significantly affect parameter

[1]Other potential applications include networks (e.g. Fowler and Christakis (2008)) and mobile telephony (e.g. Kamenica, Mullanaithan and Thaler (2011) data that have not yet been extended to real-time settings. Other emerging research areas such as precision medicine are similarly relevant (see Stern, Alexander, and Chandra, 2017).

estimates (Zwane et al, 2011). This issue is therefore relevant for real-time settings where measurement may constantly or nearly constantly occur. We provide simple measurement approaches for real-time experimentation that are broadly reconciliable with datastreams for relevant program evaluations.

We model real-time data in terms of co-existing states, of which only one is observed at a time. Real-time data may occur new data always co-exists with the old data as the datastream expands in real-time, so that parameter estimates must also be updated in real-time to miminize concept drift (e.g. Babcock et al, 2002). For program evaluations, we assume the existence of randomized treatments T_i that may vary in real-time. We also assume that the real-time assignment probabilities do not depend on the potential outcomes in the vein of the Rubin Causal Model (Rubin, (1974); Rubin (1978); also note Splawa-Neyman (1990), Imbens, Angrist and Rubin (1991); Imbens and Wooldrige (2009) and Imbens and Rubin (2015)), which we generalize to real-time settings using decision-theoretic tools (e.g. Yukalov, and D. Sornette (2014)) from human-computer interaction and computing research. Favre et al (2016) connect decision theory and experimentation in settings of uncertainty to prospect theory, which is also relevant for our approach, which focuses on real-time data. A focus on real-time measurement and data analysis may improve replicability and transparency in applied work.

The paper proceeds in the following order. Section 2 motivates the general dummy variable approach in the context of the limitations of current measurement methods. Section 3 and 4 provides the framework and approach. Sections 5 and 6 explains the generalized Rubin Causal Model and discuss datastreams before the paper ends.

2. Motivation and Generalized Dummy Variables

For a condition of interest, a dummy variable D is:

$$D = \begin{cases} 1 & \text{if the condition is met} \\ 0 & \text{if the condition is not met} \end{cases}$$

The application of dummy variables to regression analyses is often based on the following equation, where ε_i is the error term:

$$Y_i = \alpha + DX_i + \varepsilon_i$$

The classic studies of regression equations with dummy variables include Suits (1957, 1984). To avoid the so-called dummy variable trap, a single term is typically dropped from the equation for each set of dummy variables representing a categorical variable. For a review, see Wooldridge (2010).

Other equations consider the effects of two individual dummy variables D_1 and D_2 as follows:

$$Y_i = \alpha_1 + \beta_1 D_{1,i} + \beta_2 D_{2,i} + \varepsilon_i$$

Since this specification does not represent the interacted effects of D_1 and D_2, empirical presentations often yield

$$Y_i = \alpha_1 + \beta_1 D_{1,i} + \beta_2 D_{2,i} + \beta_3 (D_{1,i} D_{1,i}) + \varepsilon_i$$

I first demonstrate phenomena that violate standard dummy and interacted dummy variables and show the implied limitations of observed data.

For example, a respondent would usually state poor = 1 (to denote a poor individual who for example, lives below an economic threshold). The econometrician must assume that all poor individuals (so defined) have the same or (at least, a comparable) experience in real-time. The following simple example contradicts this assumption. Consider a data collector whose problem is to categorize dummy variables to categorize poverty based on the real-time experiences of one individual, A under a real-time scenario.

The dummy variable is unable to accomodate that the same individual may take *different* decisions when asked the *same* question at a slightly different time. Multiple interpretations of a category co-exist for each individual, but that these observations evolve in real-time so that only one experience may be observed at a time.

To further see how such situations may be common in real-time environments requires a brief discussion of data processing. Most econometrics and applied work is based on batch data processing, an efficient way of processing high volumes of data is where a group of data transactions is collected over a period of time. Data is collected, entered, processed and then the batch results are produced *before* a dummy variable is computed in the standard way by the econometrician in this context. In contrast, real time data processing involves a continual input, process and output of data. Data must be processed in real-time or in near-real-time. The values of variables may change in the interm, requiring a more general notion of dummy variables.

For the above reasons, we will isolate observed from unobserved states in real-time economic data. The effect of this observed-unobserved dichotomy exposes the role of real-time environments in exposing the observed state of poverty at each juncture is only one of multiple states that are important to the econometrician. While the obvious answer appears to simply construct two separate dummy variables and interact them to understand their collective contribution, this approach is obviously insufficient when both states *co-exist* at the level of *each* individual but *only one is observed at a time.* We also note that real-time measurement may itself affect the observed outcome (e.g. in settings where the researcher makes payments to subject during research, in this illustration). We now discuss how the creation of dummy variables can be exploited in our approach.

We provide a simplified discussion of gaining and analyzing dummy variables focusing on data collection and inference for further motivation. Consider a surveyor (agent 1) and a respondent (agent 2) who engage in a three-period relationship. (I assume that agent 2 is one of many study participants who all have their own agent 1s). In the first two periods of $t = \{1, 2, 3\}$, agent 2 can provide agent 1 with a quantity we call categorical data $c_t \in [p, q, r]$. I assume that agent 1 will receive a different data variable at $t = 3$, for an impact evaluation analysis to be performed at that third and final time period.

In this section, I briefly motivate the creation of dummy variables in the causal inference setup. I assume a single state of the world N (called a "experience") that is commonly observed by both agents although this experience or state actually happens to agent 2.

A standard dummy variable from the data is simply $w = \{0, 1\}$, and w may be categorized as 1 based on p or q or r which we assume to be mutually exclusive data entries, so that agent 1 enters "zero" for the column sections in the traditional way. Our dummy variable is $w(N)$, implying that the dummy variable is a function of the single state of the world.

For the impact evaluations at $t = 3$, we obtain the stanard Average Treatment Effects:

$$E\left[\left(Y_i\left(\left(w\left(N\right)=1\right)\right)-Y_i\left(w\left(N\right)=0\right)\right)\right]$$

I provide the simplest motivation for our framework on classifying dummy variables given different states of the world (see Section 3 for the main framework). I assume two states of the world, $N = \{x, y\}$, where N is a random type commonly observed at $t = 1, 2$.

Fact. *Different states of the world can yield different dummy variables in real-time:*

$$\left(w\left(N_x\right)\right)_{t=1} \neq \left(w\left(N_y\right)\right)_{t=1} \text{ and } \left(w\left(N_x\right)\right)_{t=2} \neq \left(w\left(N_y\right)\right)_{t=2}$$

The different states of the world and different dummy variables imply that Average Treatment Effects from the Rubin Causal Model would be similarly different. To see this, assume outcome variables at $t = 3$ to be given by $Y_i\left(w\left(N\right)\right)$. Then for our impact evaluations (based on a representative sample), the different states of the world are important, since:

$$E\left[\left(Y_i\left(\left(w\left(N_x\right)=1\right)\right)-Y_i\left(w\left(N_x\right)=0\right)\right)\right] \neq E\left[\left(Y_i\left(\left(w\left(N_y\right)=1\right)\right)-Y_i\left(w\left(N_y\right)=0\right)\right)\right]$$

If both states exist *simultaneously* in the same individual, but *only one is observed at a time*, then standard dummy variables, would be in sufficient to adequately capture the experience of interest. Analyses that consider the observed state as representative commit what we shall call a *state selection bias.* We define a more general notation for real-time scenarios, building off the standard approach while obviously benefitting from it.

Definition. (State Selection Bias): A *state selection bias* occurs when an observed state is not representative, but co-exists with other states that are not observed at a point in time. The state selection bias occurs when a single state is considered representative in real-time data and is defined as the difference between unobserved and observed states in such real-time data.

3. Social Superposition

Superposition, the phenomenon of having *theoretical* co-existing states is a foundational contribution in applied mathematics, computer science and engineering. Dirac (1947) contains mathematical foundations and notation. For relevant recent discussions in computing, see Hacohen-Gourgy et al (2016). I call my approach social superposition as I focus on the measurement consequences of co-existing multiple states in empirical data, to accomodate real-time data in economic and social science. Our discussion on multiple social states is therefore empirically-based in social and economic science applications.

I present the main framework and define the social superposition principle. Consider a social indicator variable w and number of states N. Instead of state-dependence in variables, I consider state-codependence or state-relational indicator variables. By state-codependence, we mean that a variable exists in multiple finite states at a point in time, which we shall call *positions*. A variable in a position is called a *configuration*. Suppose that a variable may have any position, so that there are different configurations which have any value of the variable w. I denote a configuration as $\mid w\rangle$.

Most General State. The most general state is given by the linear combination:

$$w_0 \mid 0\rangle + w_1 \mid 1\rangle + ... + w_N \mid N\rangle$$

Social Superposition Principle. The social superposition principle states that a combination of solutions to a linear solution is itself a solution of the linear equation in question. Therefore, there exist arbitrary superposition states of all positions.

For discrete variable w, and combination of solutions Θ, I have the following most general state:

$$\sum_{w} \Theta(w) \mid w\rangle$$

The superposition of a state with itself is the same state, unless $(w_0 + w_1) = 0$. To see this, note that $w_0 \mid X\rangle + w_1 \mid X\rangle = (w_0 + w_1) \mid X\rangle \Longrightarrow \mid X\rangle$. However, $(w_0 + w_1) = 0$ implies no state. We consider no state to be an analogy for missing data. For simplicity, we will benignly assume that$(w_0 + w_1) > 0$, so that the superposition of state with itself is the same state.

4. General Dummy Variables

In this section, I describe how to isolate mutually exclusive categories under superposition, using the social superposition theory and principle. Assume the most general state is given by $w_0 \mid 0\rangle + ... + w_N \mid N\rangle$. Each configuration $w_{n \in N} \mid n \in N\rangle$ of this most general state is mutually exclusive with the other configurations. For this reason of being mutually exclusive, each configuration is what I call a dummy variable under social superposition or a super dummy variable.

Definition 1. *A Super Dummy Variable $w_{n \in N} \mid n \in N\rangle$ is a mutually exclusive configuration under social superposition.*

A configuration of a most general state or linear combination is a super dummy variable. If a linear combination consists of only one term, the configuration is not a super dummy variable as there is no counterfactual to the term in that linear combination. This is in step with standard dummy variables which require a zero counterfactual. If a linear combinations consists of only one term, it is not a super dummy variable as there is no counterfactual to the term in that linear combination.

A basic dummy variable is a special case of social superposition, where states of interest are constrained to a linear superposition of $\mid 0\rangle$ and $\mid 1\rangle$. To show this, let $\mid 0\rangle$ and $\mid 1\rangle$ be the states that will always give 0 and 1 respectively.

A basic dummy variable is a special case of social superposition, where states of interest are constrained to a linear superposition of $\mid 0\rangle$ and $\mid 1\rangle$. To show this, let $\mid 0\rangle$ and $\mid 1\rangle$ be the states that will always give 0 and 1 respectively. Figure 1 represents both standard and super dummy variables below in a Bloch sphere. A super dummy variable is written as a superposition of $\mid 0\rangle$ and $\mid 1\rangle$. To illustrate this, note that the state $\mid 0\rangle$ is at the North Pole of the Bloch sphere, while $\mid 1\rangle$ is at the South Pole, so that an equal superposition exists somewhere on the equator of the Bloch sphere. Since the total probability of the system must be one from the social superposition principle, we can let $\mid \psi\rangle = \cos\frac{\Phi}{2} \mid 0\rangle + \sin\frac{\Phi}{2} e^{i\theta} \mid 1\rangle$, where Φ and θ are real parameters, whereby $0 \leq \Phi \leq \pi$ and $0 \leq \theta \leq 2\pi$. Any spot in the Bloch sphere would be an example of a super dummy variable.

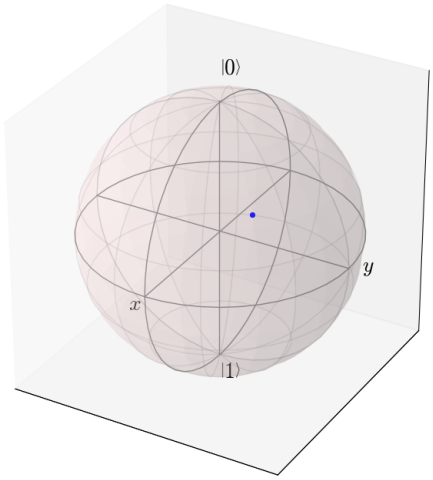

Figure 1. Bloch Sphere of Dummy Variables and Generalized Dummy Variables

We provide an analogue to interaction indicator terms used to study and create standard indicator variables from multiple indicators. We call these super interactions. Linear combinations of configurations are super interactions if the prospective combinations are mutually exclusive with other combinations. The most general state is always analogous to the sample space of a standard dummy variable, which includes present and absent states.

For simplicity, we focus on two variables which are in two different positions. For example, a state where one variable is at position a and the other is at position b is denoted $\mid a, b\rangle$.

Under social superposition, a pair of variables can be in any combination of pairs of positions Q. Recall that each term in the linear combination is a super dummy variable, where the most general state is a superposition of the possibilities:

$$\sum_{ab} Q\left(a, b\right) \mid a, b\rangle$$

An example of a super interaction is $Q\left(a, b\right) \mid a, b\rangle$, so that each configuration Q behaves as a super dummy.

Generalizing to multiple terms yields $Q\left(a, b, ...\right) \mid a, b, ...\rangle$whereby the most general state is $\sum_{ab...} Q\left(a, b, ...\right) \mid a, b, ..\rangle$

We now motivate observable data expanding on the previous discussion. The mathematical foundations in this section draws on Halmos (1987).

All observational data has the following:

(1) social positions,

(2) probabilities of social positions,

(3) measured social values,

and

(4) times of observation.

Collectively, this is to say that every datapoint has a discrete *eigenstate*. The measured value, or eigenstate of position is an *eigenvalue*. I use these points to motivate natural generalizations of discrete and continuous variables as simply sets of eigenstates. According

to Halmos (1987), a basis is a set of linearly independent vectors that represents every vector in a given vector space in a linear combination.

Fact. *Observational data has discrete eigenstates and corresponding distinct eigenvalues.*

Definition. Discrete Spectrums of Observational Data

A discrete spectrum is a variable that can vary infinitely in a continuum while only taking distinct values.

Definition. (Inner product of states). The inner product of two states ϕ and Γ is denoted $\langle \phi \mid \Gamma \rangle = \mid \phi \rangle \cdot \mid \Gamma \rangle$.

Consider an observable $\widehat{v}$, assuming the eigenstates of $\widehat{v}$ form a complete basis. .Let $\widehat{v}$ have discrete eigenstates

$$\mid 1\rangle, ..., \mid n\rangle$$

so that there are corresponding distinct eigenvalues $v_1, ...v_n$. Let a system variable be considered in state $\mid \phi\rangle = a_1 \mid 1\rangle + a_1 \mid 2\rangle + ...$, where $a_1, a_2, ...$, are complex numbers. The corresponding probabilities are

$$\Pr(a_n) = \frac{\langle n \mid \phi \rangle}{\langle \phi \mid \phi \rangle} = \langle n \mid \phi \rangle = |a_n|^2$$

, where $1 = \langle \phi \mid \phi \rangle = \sum |a_i|^2 =$ the total probability of measuring all possible states.

Measurement and State Convergence. If the observed measurement is v_n, then $\mid \phi^{'} \rangle = \mid n \rangle$ so that any repeated measurements of $\widehat{v}$ yield precisely the same v. On the other hand, if a measurement involves discrete eigenvalues and there is a discontinuous change in state, the initial superposition of several eigenstates or positions appears to converge to a single eigenstate. Being unaware of the state convergence process may lead to measurements that

are not replicable in real-time data. The observation on state experience convergence is considered the observable data, although it is not representative, given alternative states as noted in the case of dummy variables. Observable data may be biased if the experience states are not comprehensively considered.

Definition. Continuous Spectrums of Observational Data

A continuous spectrum is a set of values represented as the interval of real numbers.

I briefly re-consider $\widehat{v}$, in the context of continuous variables. Let v have cointinuous eigenstate $\mid v\rangle$, so that the eigenvalue x, filling the interval (a,b), as a continuous spectrum. For a system considered in state $\mid \phi\rangle$,

$$\mid \phi\rangle = \int_a^b a\left(v\right) \mid v\rangle dv$$

where the observable $\widehat{v}$ form an eigenbasis.

Normalization Assumption: We assume that $\int_a^b \langle \phi \mid \phi\rangle dv = 1 \; \forall \mid \phi\rangle$.

If I let $(p,q) \supseteq (r,s)$, the corresponding probability function is

$$\Pr\left(r < v < s\right) = \frac{\int_r^s \mid \langle v \mid \phi\rangle \mid^2}{\int_p^q \langle \phi \mid \phi\rangle} = \frac{\int_r^s \mid a\left(v\right) \mid^2}{\int_p^q \mid a\left(v\right) \mid^2} = \int_r^s \mid a\left(v\right) \mid^2 dv$$

Experiment Example Revisited. State selection bias can be overcome by acknowledging decision states. We interpret the above example of measurement and evaluations within real-time contexts now. We consider a lottery to be a set of outcomes and their probabilities of occurrence: $L_j = \{x_n, p_j\left(x_n\right) : n = 1 \ldots N\}$,so that $\{x_n\}$ is a set of payoffs and $p_j \in [0,1]$ and is normalized to one. Subjects can choose between two lotteries, L_1 and L_2. There exist two observables θ and ϑ, represented by operators $\hat{\theta}$and$\hat{\vartheta}$. Each observable can take two values $\theta = \{\theta_1, \theta_2\}$and $\vartheta = \{\vartheta_1, \vartheta_2\}$. We let θ_1 and θ_2 refer to the two options given to the subjects. When the lottery L_j is chosen, $\theta_j \left(j = 1,2\right)$ and ϑ is a measure of the relevant uncertainty that exists in real-time. Examples may be the lack of confidence on the part of the subject that the evaluation is properly understood, or whether

they believe that the lottery will proceed as explained by the evaluator. If this changes in real-time, this real-time uncertainty, may manifest as a change in the real-time data. The eigenstates of each operator form an orthonormal basis of the relevant Hilbert space, so that $\mathcal{H}_\theta = \text{span}\{|\,\theta_1\rangle, |\,\theta_2\rangle\}$ and $\mathcal{H}_\vartheta = \text{span}\{|\,\vartheta_1\rangle, |\,\vartheta_2\rangle\}$. The subject decision maker's state is represented as $\mathcal{H}_{AB} = \text{span}\{|\,\theta_1\vartheta_1\rangle, |\,\theta_1\vartheta_2\rangle, |\,\theta_2\vartheta_1\rangle, \theta_2\vartheta_2\rangle\}$, which fully represents the relevant states. We now provide the general Rubin Causal Framework which draws on the above arguments.

5. Superposition and the Rubin Causal Framework

We follow Imbens and Wooldridge (2009). Assume $N \in i = 1, ..., N$. We let $\boldsymbol{W}$ refer to an $N-$vector of individuals, where some individuals were randomly assigned to the program under evaluation, while others were not, so that N_0 and N_1 are control and treated units respectively. We observe a $K-$dimensional column vector of covariates X_i so that $\boldsymbol{X}$ refers to a $N \times K$ matrix with i-th row equal X'.

The status of an individual is:

$$W_i = \begin{cases} W_i = 1 & \text{if individual } i \text{ was treated in the program} \\ W_i = 0 & \text{if individual } i\text{was not treated in the program} \end{cases}$$

In addition to assuming random states, we relax standard assumptions of temporal stability, causal transience, unit homogeneity, independence and constant effects (see Holland (1986) for details on these basic RCM assumptions). The new assumptions are as follows.

Assumption. *Random states.*

I assume that experiences or states are randomly distributed.

Assumption. *Temporal Instability.*

I relax the assumption of constant response over time. We assume inconstant responses over time, as the experiential differences in treatment exposure compensate for inconstant responses. From the basic example (Section 2), this implies that $(w(N_x))_{t=1} \neq (w(N_x))_{t=2}$.

Assumption. *Causal Intransience.*

I relax the presumption of causal transience. We assume that the effect of the cause and the measurement process that result in Y_i is intransient and does change W_i enough to affect $Y_i(W_i)$ measured later. Research by Zwane et al (2011) found that being surveyed in of itself can affect behavior, and causal intransience may be one relevant interpretation. For simplicity, I assume that the experiential aspect of treatment exposure cancels out the effect of the measurement process in this section.

Assumption. *Unit Heterogeneity.*

I relax unit homogeneity. I assume that the outcomes $Y_{i=t,c}$ (t = treatment; c = control) for two units u_1 and u_2 are not equal, so that $Y_t(u_1) \neq Y_t(u_2)$, based on the experiences or states.

Assumption. *Weak Independence.*

I do not require complete independence, given the importance of random experiences. I assume a weak form of independence, so that $E(Y_{W_i=1}) \neq E(Y_{W_i=1}|W_i = 1)$and $E(Y_{W_i=0}) \neq E(Y_{W_i=0}|W_i = 0)$. On the other hand, one may first assume non-random states, which would require a complete independence assumption for our Rubin Causal Model generalization.

Assumption. *Inconstant Effects.*

The impact of $W_i = 1$ on every unit is not assumed to be the same.

Conjecture. *The treatment $W_i = 1$ and the control $W_i = 0$ respectively exist in multiple treatment states and control states at the same time, and for both the treatment and control, only one treatment and control state is observed at a single point in time.*

This argument follows from social superposition. The Rubin Causal Model under social superposition is:

$$Y_i = Y_i(W_i) = Y_i(W_i = 0 \mid 1\rangle + ... + W_i = 0 \mid N_0\rangle\rangle) + Y_i(W_i = 1 \mid 1\rangle + ... + W_i = 1 \mid N_1\rangle\rangle)$$

Proposition. *Average Treatment Effects under Social Superposition: True Average State Treatment Effects (TASTE)*

The Average Treatment Effect under social superposition is what we shall call the True Average State Treatment Effects (TASTE) is provided by the formula

$$E\left[Y_i\left(W_i = 1 \mid 1\rangle + ... + W_i = 1 \mid N_1\rangle\right) - Y_i\left(W_i = 0 \mid 1\rangle + ... + W_i = 0 \mid N_0\rangle\right)\right]$$

.

The TASTE becomes:

$$\left\{\frac{\sum\left(W_i = 1 \mid 1\rangle + ... + W_i = 1 \mid N_1\rangle\right) Y_i}{\sum\left(W_i = 1 \mid 1\rangle + ... + W_i = 1 \mid N_1\rangle\right)}\right\} - \left\{\frac{\sum Y_i\left(W_i = 0 \mid 1\rangle + ... + W_i = 0 \mid N_0\rangle\right) Y_i}{\sum Y_i\left(W_i = 0 \mid 1\rangle + ... + W_i = 0 \mid N_0\rangle\right)}\right\}$$

Corollary. *TASTE Estimates Are Externally Valid.*

Proof. This proposition is motivated by intuition. Since the treatments and controls account for the full space of experience states in the population, *TASTE estimates are externally valid.* To see this in the case of the treatment experiences, note that if estimates were not externally valid, there would exist a state not controlled for in the space of experience states, which is contradicted by the definition of the most general state. The same argument holds for control experiences.

□

To see that the original Rubin Causal Model represents a special case, note that the potential outcomes result provides the existence of two potential outcomes, $Y_i(0)$and $Y_i(1)$, which reflect outcomes if $W_i = 1$ or $W_i = 0$ respectively:

$$= \begin{cases} Y_i\left(\mid 0\rangle\right) = Y_i(0) & \text{if } W_i = 0 \\ Y_i\left(\mid 1\rangle\right) = Y_i(1) & \text{if } W_i = 1 \end{cases}$$

Following the relationship between standard and super dummy variables, the RCM is a special case whereby where we are constrained to a linear superposition of $\mid 0\rangle$ and $\mid 1\rangle$for W_i. The potential outcomes are $Y_i\left(\mid 0\rangle\right) = Y_i(0)$and $Y_i\left(\mid 1\rangle\right) = Y_i(1)$ respectively.

Given that the act of high-dimensional measurement may itself affect observations in real-time, we transition to the use of datastreams, digitally encoded coherent signal sequences used to transmit or receive information that is in the process of being transmitted.

6. Real-Time Datastreams

Datastreams can be thought of as generalizations of datasets. These are a sequence of digitally encoded coherent signals (packets of data) used to transmit or receive information that is *in the process of being transmitted.* We consider the simple case where each new data entry replaces and old entry and the econometrician must compute descriptive statistics before considering the case of performing evaluations where data entries co-exist. To illustrate, a health scientist may generalize the analysis of Green and MacDonald (1981) on the static effect of intravenous glucose on body temperature to real-time data by asking, "what are the effects of intravenous glucose T on body temperature outcome Y for treatment patients relative to control patients in real-time?"

Real-Time Measurements and Descriptive Statistics. An focus on real-time datastreams requires detailed statistical definitions on time. These definitions are in the vein of Zhu and Shasha (2002). A *timepoint* is the smallest unit of time over which data collection occurs (e.g. a second). A *basic window* is a consecutive subsequence over which a digest of data is maintained incrementally (e.g. a few minutes). A *sliding window* is a user-defined sub-sequence of basic windows over which the user requires statistics (e.g. a hour).

Assuming a length of a sliding window w and current timepoint t, for a specific sliding window q, we compute $stat\left(\mathcal{D}\left(X_i, T_i, Y_i\right), q\left(k\right)\right)$in the subsequence $\mathcal{D}\left(X_i, T_i, Y_i\right)[t-w+1..t]$.

Consider the stream $\mathcal{D}_i, i = 1, \ldots, w$. We will assume that we are in a position to monitor the following descriptive statistics in the real-time data, such as single stream statistics (average, standard deviations), real-time correlation coefficients $corr\left(\mathcal{X}, \mathcal{Y}\right)$, autocorrelation (the correlation of the series with its previous version), the sensitivity of a stream $\mathcal{T}$ to the

values of another stream $\mathcal{Y}$ (or weighted collection of streams), or$beta\left(\mathcal{X},\mathcal{Y}\right)$, and others as necessary.

To compute statistics over a sliding window, we maintain a synopsis data structure for the stream to compute the statistics in a timely manner. Our framework subdivides the sliding windows equally into shorter basic windows, to facilitate the efficient elimination of old data and the incorporation of new data. Digests are kept for both basic and sliding windows.

Let w be the size of a sliding window, b be the size of a basic window, k be the number of basic windows within a sliding window. The data within a sliding window is $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[t-w+1..t\right]$. From the above, $w=kb$.

Let $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[0\right],...,\mathcal{D}\left(X_i,T_i,Y_i\right)\left[k-1\right]$ represent a sequence of basic windows, whereby

$$\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i\right]=\mathcal{D}\left(X_i,T_i,Y_i\right)\left[\left(t-w\right)+ib+1..\left(t-w\right)+\left(i+1\right)b\right]$$

The new basic window is $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[k\right]$ and $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[0\right]$ is the expiring basic window. The $j-$th value in the basic window $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i\right]$ is $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i;j\right]$.

The size of the basic window is important since in some cases the researcher may report all statistics for basic window i before basic window $i+1$ completes (at which point the researcher must compute statistics for window $i+1$.

The moving average and moving standard deviations are as follows. The information to be maintained for the moving average is

$$\sum\left(\mathcal{D}\left(X_i,T_i,Y_i\right)\left[t-w+1\ldots t\right]\right).$$

For each basic window $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i\right]$, we maintain the digest

$$\sum\left(\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i\right]\right)=\sum_{j=1}^{b}\mathcal{D}\left(X_i,T_i,Y_i\right)\left[i;\quad j\right].$$

After b new datapoints become available, we compute the sum over the new basic window $\mathcal{D}\left(X_i,T_i,Y_i\right)\left[k\right]$. The sum over the sliding window may be updated as:

$$\sum_{new} \left(\mathcal{D}\left(X_i, T_i, Y_i\right)\right) = \sum_{old} \left(\mathcal{D}\left(X_i, T_i, Y_i\right)\right)$$

$$+ \sum \mathcal{D}\left(X_i, T_i, Y_i\right)[k] - \sum \mathcal{D}\left(X_i, T_i, Y_i\right)[0]$$

Correlational Statistics in Real-Time: Assume a series of covariates $\boldsymbol{s}^x$ as well as a series of error terms, $\boldsymbol{s}^\epsilon$. For causality, we assume strict exogeneity of the treatment series, so that $\boldsymbol{E}\left[\boldsymbol{s}^\epsilon | \boldsymbol{s}^x\right] = \mathbf{0}$ where $s^x = [\alpha, s^t]$, which is based on an assumed random assignment of T_i and of the treatment series s^t, which is implemented in a randomized manner and arrives at random. We will use covariate balance in real-time to verify the exogeneity of treatment streams throughout the real-time data.

Theorem 1: The True Average State Treatment Effects (TASTE) are estimable in real-time.

Proof. We now show how treatment effects may be studied in real-time data using datastreams where old and new data entries co-exist in real-time. Let $\mathcal{D}_i$ or $D[i]$ denote the value of datastream $\mathcal{D}$ at timepoint i. For timepoints i through j inclusive, the corresponding stream is $\mathcal{D}\left\{(X_i, T_i, Y_i)\right\}_{i=1}^{n}[i...j]$. The stream is summarized as $\mathcal{D}[i...j]$. A stream with stream identifier i is denoted $\left[\mathcal{D}\left\{(X_i, T_i, Y_i)\right\}_{i=1}^{n}\right]^i$.

Proof. We let the real-time data be represented in the following form:

$$\mathcal{D} = \{(X_i, T_i, Y_i)\}$$

which are independent identically distributed triples (X_i, T_i, Y_i)

□

We let $X \in \mathbb{R}^p$ represent the subject characteristics measured at baseline; $T \in \mathcal{T}$ refers to the binary or continuous treatment received by the subjects; and the outcome is summarized by $Y \in \mathbb{R}$. The real-time data is generated according to the linear form $Y_i = \alpha + \beta T_i + \epsilon_i$ for a constant α and standard error term ϵ_i.

For a window of interest $[p, q]$, we represent the statistics of interest by

$$stat\left(\mathcal{D}_j^{i_1}, \mathcal{D}_j^{i_2}, ..., \mathcal{D}_j^{i_k}, j \in [p,q]\right)$$

or

$$stat\left(\left(\mathcal{D}\left\{(X_i, T_i, Y_i)\right\}_{i=1}^n\right)_j^{i_1}, \left(\mathcal{D}\left\{(X_i, T_i, Y_i)\right\}_{i=1}^n\right)_j^{i_2}, ..., \left(\mathcal{D}\left\{(X_i, T_i, Y_i)\right\}_{i=1}^n\right)_j^{i_k}\right)$$

A treatment rule is the map $\pi : \mathbb{R}^p \longrightarrow \mathcal{T}$ so that a subject observed with $X = x$ receives treatment $\pi(x)$. The potential outcome under treatment $t \in \mathcal{T}$ is $Y^*(t)$so that the potential outcome under π is defined as $Y^*(\pi) = Y^*\{\pi(X)\}$. The marginal mean outcome is $E\{Y^*(\pi)\}$ and the conditional density of T given X is $p(t|X)$. These definitions are standard.

We generate the real-time treatment effects as follows. The average treatment effect in real time over $[p,q]$ is given by the difference between the expectations of real-time outcomes attributable to treatment t_1 and the expectations of real-time outcomes attributable to control t_0 over the same window.

$$\mathbb{E}\left(\mathcal{Y}(t_1)_j^{i_1}, \mathcal{Y}(t_1)_j^{i_2}, ..., \mathcal{Y}(t_1)_j^{i_k}, j \in [p,q]\right)$$

$$-\mathbb{E}\left(\mathcal{Y}(t_0)_j^{i_1}, \mathcal{Y}(t_0)_j^{i_2}, ..., \mathcal{Y}(t_0)_j^{i_k}, j \in [p,q]\right)$$

where the real-time outcomes attributable to the treatment are documented as

$$\mathcal{Y}(t)_j^{i_1}, \mathcal{Y}(t)_j^{i_2}, ..., \mathcal{Y}(t)_j^{i_k}$$

and the real-time outcomes attributable to the control are given by

$$\mathcal{Y}(c)_j^{i_1}, \mathcal{Y}(c)_j^{i_2}, ..., \mathcal{Y}(c)_j^{i_k}.$$

If the above can be computed, we identify this formula as the Rubin Causal Model in real-time. Unit-level real-time causal effects are

$$\mathbb{E}\{\left(\mathcal{Y}\left(t_1\right)_j^{i_1}, \mathcal{Y}\left(t_1\right)_j^{i_2}, ..., \mathcal{Y}\left(t_1\right)_j^{i_k}\right)$$

$$- \left(\mathcal{Y}\left(t_0\right)_j^{i_1}, \mathcal{Y}\left(t_0\right)_j^{i_2}, ..., \mathcal{Y}\left(t_0\right)_j^{i_k}, \right)\}$$

for $j \in [p, q]$. The real-time Population Average Treatment effect on the treated is defined by

$$\mathbb{E}\{\left(\mathcal{Y}\left(t_1\right)_j^{i_1}, \mathcal{Y}\left(t_1\right)_j^{i_2}, ..., \mathcal{Y}\left(t_1\right)_j^{i_k}\right)$$

$$- \left(\mathcal{Y}\left(t_0\right)_j^{i_1}, \mathcal{Y}\left(t_0\right)_j^{i_2}, ..., \mathcal{Y}\left(t_0\right)_j^{i_k}\right)\}|\left(t_0\right)_j^{i_1}, \left(t_0\right)_j^{i_2}, ..., \left(t_0\right)_j^{i_k}$$

for $j \in [p, q]$ $\square$.

Covariate Balance in Real-Time: We discuss cases where causal inference is based on background variables such as sociodemographic indicators or pure experiments, based on the standard assumptions of the Rubin Causal Model. Hansen and Bowers (2008) use Fisher's randomization inference when confirming covariate balance in static data, which may be done with stratification (e.g. Jin and Rubin 2008) to ensure covariate balance. We focus this approach of discussing identification on real-time settings.

Simple Cluster-based Randomization: The index $i_k = (1, \ldots n)_k$ refers to assignment units for each window k. The treatment assignment of the ith cluster of observation units at the k-th window is z_{ik}. The total number of x-values for observation units in cluster i is x_{ik}. The size of the cluster is c_{ik}

The observed difference between treatment and control groups in window k is $d_{pk}\left(\boldsymbol{z},\boldsymbol{x}\right)$, so that:

$$d_{pk}\left(\boldsymbol{z}, \boldsymbol{x}\right) = \left[\frac{\boldsymbol{z_k}^{\boldsymbol{t}}\boldsymbol{x_k}}{\boldsymbol{z_k^t c_k}}\right] - \left[\frac{\left(\boldsymbol{1 - z_k}\right)^{\boldsymbol{t}}\boldsymbol{x_k}}{\left(\boldsymbol{1 - z_k}\right)^{t}\boldsymbol{c_k}}\right]$$

and $d_{pk}\left(\boldsymbol{z}, \boldsymbol{x}\right)$ refers to the difference in treatment group averages and control group averages at each window k. If A_k is the set of treatment assignments from which the actual

assignment $\boldsymbol{z_k}$ is selected, we have the following two-sided randomization $p-$value attaching to the hypothesis of nonselection on x as:

$$= \boldsymbol{P}\left(\left|d_{pk}\left(\boldsymbol{Z_k}, \boldsymbol{x_k}\right)\right| > \left|d_{pk}\left(\boldsymbol{Z_k}, \boldsymbol{x_k}\right)\right|\right) + \frac{1}{2}\boldsymbol{P}\left(\left|d_{pk}\left(\boldsymbol{Z_k}, \boldsymbol{x_k}\right)\right| = \left|d_{pk}\left(\boldsymbol{Z_k}, \boldsymbol{x_k}\right)\right|\right)$$

The random vector $\boldsymbol{Z_k}$ is uniformly distributed on possible treatment assignments A_k for all k. If simple random samples of n_{tb} are selected from n_b clusters for a treatment stream within blocks b, with the rest assigned to a control stream, we have:

$$d_k\left(\boldsymbol{z}, \boldsymbol{x}\right) = \left\{\left(\frac{1}{\sum h_b \bar{c}_b}\right)\left[\sum_{b=1}^{B} \boldsymbol{Z}_b^t \boldsymbol{x}_b - \sum_{b=1}^{B} n_{tb}\left(\frac{\mathbf{1}^t \boldsymbol{x}_b}{n_b}\right)\right]\right\}_k$$

$$\mathbb{E}\left(d_k\left(\boldsymbol{Z}, \boldsymbol{x}\right)\right) = \mathbb{E}\left(d_k\left(\boldsymbol{Z}, \boldsymbol{v}\right)\right) = 0,$$

for all k.

The variance and covariance are given per window as follows:

$$\mathrm{Var}\left(d\left(\boldsymbol{Z}, \boldsymbol{x}\right)\right) = \left\{\left(\frac{1}{\sum h_b \bar{c}_b}\right)^2 \sum_{b=1}^{B} h_b \bar{c}_b \frac{s^2\left(\boldsymbol{x}_b\right)}{\bar{c}_b}\right\}_k$$

$$\mathrm{Cov}\left(d_k\left(\boldsymbol{Z}, \boldsymbol{x}\right), \left(d_k\left(\boldsymbol{Z}, \boldsymbol{v}\right)\right)\right) = \left\{\left(\frac{1}{\sum h_b \bar{c}_b}\right)^2 \sum_{b=1}^{B} h_b \bar{c}_b \frac{s_k\left(\boldsymbol{x}_b; \boldsymbol{v}_b\right)}{\bar{c}_b}\right\}_k$$

for all k.

Simple Within-Block Randomization: Consider $b \in B$ strata, from which simple random samples of $n_{t1} \ldots n_{tB}$ clusters are selected into the treatment group from $n_1 \ldots n_B$ total clusters. As a random variable, the adjusted difference of treatment and control averages at every window k is given as:

$$d_k\left(\boldsymbol{Z}, \boldsymbol{x}\right) = \left\{\sum_{b=1}^{B} w_b \left[\frac{\boldsymbol{Z}_b^t \boldsymbol{x}_b}{c_{tb}} - \frac{\left(\mathbf{1} - \boldsymbol{Z}_b^t \boldsymbol{x}_b\right)}{\left(c - c_{tb}\right)}\right]\right\}_k$$

$$= \left\{\sum_{b=1}^{B} w_b h_b^{-1} \bar{c}_b^{-1} \boldsymbol{Z}_b^t \boldsymbol{x}_b - \sum_{b=1}^{B} w_b c_b^{-1}\left(n_b - n_{tb}\right)^{-1} \mathbf{1}_b^t \boldsymbol{x}_b\right\}_k$$

7. Regressions with Exogenous Real-Time Experimental Treatments

Guided by the average treatment effects in the first section, we start with the Ordinary Least Square regressions, assuming exogenous treatments. However, although it is necessary to continuously update the regression model parameters while receiving new data, the tremendous data volume renders it impossible to scan the entire data stream multiple times to regenerate the regression model parameters. The Approximate Stream Regression approach approximately estimates the regression parameters by using a weighted average of the parameters obtained from current window data and the regression parameters of pervious windows.

We now assume the existence of *concept drift:* meaning that regression coefficients can change over time in real-time data. The Approximate Stream Regression adapts to the incidence of concept drift, automatically updating regression coefficients to current values without having to rebuild the entire model during a program evaluation.

Assuming the model with the following functional form:

$$Y = \beta T + \varepsilon$$

If we let β_{k-1} refer to the coefficient over window $k-1$, then according to the Exponential Weighted Moving Average (Brown and Meyer, 1961), we obtain the following,

$$\beta_k = (1-\sigma)\,\beta'_k + \sigma\beta_{k-1}$$

such that σ is a smoothing factor and constant value where both $\sigma \leq 1$ and $(1-\sigma) \leq 1$. Also, β'_k is a parameter vector for the kth window (calculated considering only the data records of the kth window). Also, β_{k-1} is the parameter vector for the $(k-1)$th window (calculated considering all data records seen up to $(k-1)$th window). Also, β_k is the refined

parameter vector for the kth window calculated considering all data records seen up to the kth window. Substituting and expanding the Right Hand Side,

$$Y_k = \{(1-\sigma)\,\beta_k' + \sigma\beta_{k-1}\}\,T_k + \varepsilon_k$$

$$Y_k = \{(1-\sigma)\,\beta_k'\}\,T_k + \{\sigma\beta_{k-1}\}\,T_k + \varepsilon_k$$

$$Y_k = \left\{(1-\sigma)\,\beta_k' + \cdots + (1-\sigma)\,(\sigma)^j\,\beta_{k-j} + (\sigma)^k\,\beta_1\right\}T_k + \varepsilon_k$$

$$Y_k = (1-\sigma)\,\beta_k' T_k + \cdots + (1-\sigma)\,(\sigma)^j\,\beta_{k-j} T_k + (\sigma)^k\,\beta_1 T_k + \varepsilon_k$$

Assume exogeneity in real-time; $cov\,(\varepsilon,(T)) = 0$

Theorem 2: The bias of the simple experimental coefficient when used in real-time environments where concept drift is relevant is given by:

$$\left\{(1-\sigma)\,\beta_k' + \cdots + (1-\sigma)\,(\sigma)^j\,\beta_{k-j} + (\sigma)^k\,\beta_1\right\} - \left\{\left(T_k'T_k\right)^{-1}T_k'Y_k\right\} > 0$$

Proof. The expression is always positive due to the existence of concept drift. Since time is strictly increasing, concept drift is always positive which leads to larger estimates in real-time data. The bias is the difference in estimates.

□

Difference-in-difference estimators are used to compute treatment effects when time trends have the potential to confound treatment impacts (see Imbens and Wooldridge, 2009, also see Bertrand, Duflo, and Mullainathan 2004 for clustered standard errors). We provide a real-time version of difference-in-difference estimation. Given the importance of time factors

and concept drift, we use the Approximate Stream Regression approach. The difference-in-difference estimator produces causal effects, assuming that parallel trend assumptions hold.

$$Y_k = \alpha_k + \beta_k T_k + \gamma_k t_k + \delta_k \left(T_k \cdot t_k\right) + \varepsilon_k, \text{ for all } k.$$

α_k : constant term for window k

β_k : treatment group specific effect for window k, which accounts for average permanent differences between treatment and control

γ_k : time trend (in real-time) common to treatment and control groups in window k

δ_k : true effect of treatment in window k

We make the following points:

We are interested in an unbiased estimate $\hat{\delta}$, whereby $E\left[\hat{\delta}\right] = \delta$. The error term is zero on average: $E\left[\varepsilon_k\right] = 0$. The error term is not correlated with the other variables (exogeneity): $cov\left(\varepsilon_k, (T_k)\right) = 0$ and $cov\left(\varepsilon_k, (t_k)\right) = 0$.

Parallel trend assumption: $cov\left(\varepsilon_k, (T_k \cdot t_k)\right) = 0$. This identifying assumption is based on the premise that the real-time average change in the control group in real-time represents the counterfactual change in the treatment group if there were no treatment exposure received by the real-time data. We consider the average change in real-time outcomes for the treated in the absence of treatment to equal the average change in outcome for the non-treated.

From the difference-in-difference regression equation, also note that

$$E\left[Y_0^T\right] = \alpha + \beta$$

$$E\left[Y_1^T\right] = \alpha + \beta + \gamma + \delta$$

$$E\left[Y_0^C\right] = \alpha$$

$$E\left[Y_1^C\right] = \alpha + \gamma$$

From the Exponential Weighted Moving Average, we obtain the following coefficients for the difference-in-difference equation:

$$\delta_k = (1-\theta)\,\delta'_k + \theta\delta_{k-1}$$

whereby θ is a smoothing factor and constant value where both $\theta \leq 1$ and $(1-\theta) \leq 1$. Also, δ'_k is a parameter vector for the kth window (calculated considering only the data records of the kth window). Also, δ_{k-1} is the parameter vector for the $(k-1)$th window (calculated considering all data records seen up to $(k-1)$th window). Also, δ_k is the refined parameter vector for the kth window calculated considering all data records seen up to the kth window. We similarly define the following:

$$\gamma_k = (1-\vartheta)\,\gamma'_k + \vartheta\gamma_{k-1}$$

and

$$\beta_k = (1-\tau)\,\beta'_k + \tau\beta_{k-1}$$

Expanding

$$\delta_k = (1-\theta)\,\delta'_k + \cdots + (1-\theta)\,(\theta)^j\,\delta_{k-j} + (\theta)^k\,\delta_1$$

$$\gamma_k = (1-\vartheta)\,\gamma'_k + \cdots + (1-\vartheta)\,(\vartheta)^j\,\gamma_{k-j} + (\vartheta)^k\,\gamma_1$$

$$\beta_k = (1-\tau)\,\beta'_k + \cdots + (1-\tau)\,(\tau)^j\,\beta_{k-j} + (\tau)^k\,\beta_1$$

The refined value is a weighted combination of the previous values and the current value. By varying the smoothing factor τ, we can adjust the importance of the historical data. There is a negative relationship between the size of the smoothing factor and the weight

assigned to the current data, as well as a positive relationship between the size of the smoothing factor and the degree to which the historical data affects the regression parameters. For equal weights between current values and past historical data, $\tau = \frac{1}{2}$.

Depending on the characteristics of the data stream, we may specify a desired half-life for the current window. Then the smoothing constant θ may be calculated in the following manner using the following half-life window principle:

$$\theta^k = e^{-\lambda k}$$

The value of λ can be calculated by $e^{-\lambda t_{\frac{1}{2}}} = \frac{1}{2}$, whereby $t_{\frac{1}{2}}$ is the desired approximate half-life time of a particular set of records. Thus, $\lambda = \frac{\ln(2)}{t_{\frac{1}{2}}}$ and $\theta = e^{\left[\frac{\ln(2)}{t_{\frac{1}{2}}}\right]}$.

For example, if we need to reduce the impact of the records in the kth window approximately by half, when we proceed to process the $(k+5)$th window, then $t_{\frac{1}{2}} = 5$ and $e^{\left[\frac{\ln(2)}{5}\right]}$.

If concept drift is present, recent samples will better reflect the status of the data stream. For such non-statiionary data we can use a smaller half-life value, so that θ will be adjusted to asign a higher weight on recent samples and generate more accurate regression model.

The difference-in-difference regression equation is given as:

$$Y_k = \alpha_k + \left\{(1-\tau)\,\beta'_k + \cdots + (1-\tau)\,(\tau)^j\,\beta_{k-j} + (\tau)^k\,\beta_1\right\} T_k$$

$$+ \left\{(1-\vartheta)\,\gamma'_k + \cdots + (1-\vartheta)\,(\vartheta)^j\,\gamma_{k-j} + (\vartheta)^k\,\gamma_1\right\} t_k$$

$$+ \left\{(1-\theta)\,\delta'_k + \cdots + (1-\theta)\,(\theta)^j\,\delta_{k-j} + (\theta)^k\,\delta_1\right\} (T_k \cdot t_k) + \varepsilon_k, \text{ for all } k$$

Theorem 3: The real-time difference-in-difference estimator $\delta_{RTDD} = \bar{Y}^T_{1k} - \bar{Y}^T_{0k} - \left(\bar{Y}^C_{1k} - \bar{Y}^C_{0k}\right)$ is unbiased.

The proof is in the Appendix.

We provide simple algorithms to implement the approaches, proposing simple scripts to perform measurement and causal inference instantaneously without any delay in timeliness. The procedures are summarized in Algorithms 1-3. These approaches allow for the estimation of Average Causal Effects using the different data environments. To verify the randomization of treatments in practice, real-time covariate balance may be utilized, so that real-time covariates are considered balanced when they are not statistically different from one another. However, some data may be constrained in this regard (e.g. if covariate variables are not available and the treatment is randomly implemented, as assumed). The real-time covariate balance may support the real-time parallel trend observations in the difference-in-difference implementation.

Algorithm 1 General Method of Simple Estimation of Average Causal Effects in Real-Time Data with Concept Drift (ASR)

Require: $k \geq 0$
Ensure: All parameter vectors for the current window of interest and the previous window
1: **while** real-time covariate variables balance **do**
2: Solve $(Y_k = \beta_k T_k + \varepsilon_k)$
3: **end while**

Algorithm 2 General Method of Difference-in-Difference Estimation in Real-Time Data

Require: $k \geq 0$
Ensure: All parameter vectors for the current window of interest and the previous window
1: **while** parallel trends assumption holds and real-time covariate variables balance **do**
2: Solve $(Y_k = \alpha_k + \beta_k T_k + \gamma_k t_k + \delta_k (T_k \cdot t_k) + \varepsilon_k)$
3: **end while**

8. A Secondary Problem of Causal Inference in Real-Time

Economists have long been acutely aware of broadly-related measurement issues (see Morgenstern, 1950), and we now exploit the emergence of real-time data environments to contribute to this discussion on how economic measurement can complicate causal inference and program evaluations. We illustrate the role of real-time data in complicating an important concept in causal inference. To measure the causal effect of study Treatment

A versus study Treatment B, the investigator requires the outcome for the identical study individual in both alternative futures. Since it is impossible to see both potential outcomes at once, one of the potential outcomes is always missing. This observation is known as the Fundamental Problem of Causal Inference (Holland, 1986). The Fundamental Problem of Causal Inference occurs because we do not ever simultaneously observe an individual with and without a treatment (summarized in Figure 1).

Figure 2. Fundamental Problem of Causal Inference: We do not observe control and treatment simultaneously in static datasets.

Our approach based on the real-time experience of a treatment has implications for this particular inference problem. One consequence of our focus on real-time datastreams is that a treatment real-time stream may overlap with a control real-time stream as exclusively as a result of the measurement process. If this occurs, an agent may experience a randomized treatment in a defined state that gives the econometrician an impression that the subject is *not* actually experiencing the treatment during a program evaluation (i.e. as though the individual were part of the control group). To illustrate, a randomized set of individuals might be partially observed to experience a financial treatment in a way that seems exactly the same as not experiencing the treatment in the real-time stream. The opposite experience of being in a control state that seems identical to a treatment state may also occur. For example, individuals who did not receive a financial treatment may be in a state which allows them to be comparable to the treatment group for experiential reasons alone. We collectively call these phenomena the Secondary Problem of Causal Inference in Real-Time Data. This complementary Problem is merely additional motivation for general impact evaluations in real-time settings. To illustrate, the region of Figure 3 where the control

and treatment real-time sets intersect ("SP") identifies the Secondary Problem of Causal Inference.

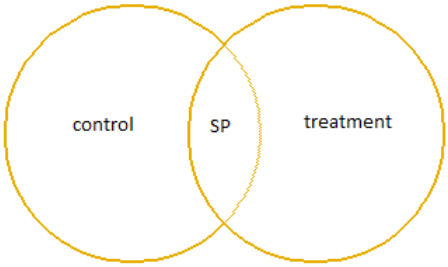

Figure 3. Secondary Problem of Causal Inference in Real-Time Data

This dilemma may be one reason why some experiments may not appear to be replicable in real-time data.

Definition. The Secondary Problem in real-time occurs in the event where

$$\left((t_1)_j^{i_k}, j \in [p,q]\right) = \left((t_0)_j^{i_k}, j \in [p,q]\right).$$

9. Conclusion

This paper focuses on experimentation in real-time data environments. In the simple case, we assumed that that no new variables emerged in the real-time data. If necessary, the introduction of new variables may also be considered and added to gain plausible treatment effects. Also real-time fixed effects and window dummies are compatible with the regressions as well as clustered standard errors. Estimation of treatment effects in observational data collected in real-time may benefit from estimated dose-response functions (e.g. Imbens, 2000), and other aspects of the program evaluation toolkit, but generalized to focus on real-time information from datastreams. Collaborations between economists and computer

scientists may be beneficial to policy makers interested in exploiting non-traditional data structures containing potential information.

Appendix: Proof of Theorem 3

To show that the real-time difference-in-difference estimator is unbiased we expand the difference-in-difference regression equation as follows,

$$Y_k = \alpha_k + \beta_k T_k + \gamma_k t_k + \delta_k \left(T_k \cdot t_k\right) + \varepsilon_k, \text{ for all } k.$$

$$Y_k = \alpha_k + \left\{(1-\tau)\,\beta_k' + \tau\beta_{k-1}\right\} T_k$$

$$+ \left\{(1-\vartheta)\,\gamma_k' + \vartheta\gamma_{k-1}\right\} t_k + \left\{(1-\theta)\,\delta_k' + \theta\delta_{k-1}\right\} \left(T_k \cdot t_k\right) + \varepsilon_k, \text{ for all } k$$

Taking expectations and substituting yields the following:

$$E\left[Y_0^T\right]_k = \alpha_k + \left\{(1-\tau)\,\beta_k' + \tau\beta_{k-1}\right\}$$

$$= \alpha_k + \left\{(1-\tau)\,\beta_k' + \cdots + (1-\tau)\,(\tau)^j\,\beta_{k-j} + (\tau)^k\,\beta_1\right\}$$

$$E\left[Y_1^T\right]_k = \alpha_k + \left\{(1-\tau)\,\beta_k' + \tau\beta_{k-1}\right\} + (1-\vartheta)\,\gamma_k' + \vartheta\gamma_{k-1} + (1-\theta)\,\delta_k' + \theta\delta_{k-1}$$

$$= \alpha_k + \left\{(1-\tau)\,\beta_k' + \cdots + (1-\tau)\,(\tau)^j\,\beta_{k-j} + (\tau)^k\,\beta_1\right\}$$

$$+ \left\{(1-\vartheta)\,\gamma_k' + \cdots + (1-\vartheta)\,(\vartheta)^j\,\gamma_{k-j} + (\vartheta)^k\,\gamma_1\right\}$$

$$+\left\{(1-\theta)\,\delta'_k+\cdots+(1-\theta)\,(\theta)^j\,\delta_{k-j}+(\theta)^k\,\delta_1\right\}$$

$$E\left[Y_0^C\right]_k=\alpha_k$$

$$E\left[Y_1^C\right]_k=\alpha_k+(1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1}$$

$$=\alpha_k+\left\{(1-\vartheta)\,\gamma'_k+\cdots+(1-\vartheta)\,(\vartheta)^j\,\gamma_{k-j}+(\vartheta)^k\,\gamma_1\right\}$$

To conclude that the difference-in-difference estimator is unbiased, substitute the expressions as follows:

$$\delta_{RTDD_k}=E\left[\bar{Y}_{1k}^T\right]-E\left[\bar{Y}_{0k}^T\right]-\left(E\left[\bar{Y}_{1k}^C\right]-E\left[\bar{Y}_{0k}^C\right]\right)$$

$$(\alpha_k+\{(1-\tau)\,\beta'_k+\tau\beta_{k-1}\}+(1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1}+(1-\theta)\,\delta'_k+\theta\delta_{k-1})$$

$$-(\alpha_k+\{(1-\tau)\,\beta'_k+\tau\beta_{k-1}\})-(\alpha_k+(1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1})$$

$$-\alpha_k+(1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1}-((1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1})$$

$$\delta_{RTDD_k}=((1-\vartheta)\,\gamma'_k+\vartheta\gamma_{k-1})=\delta_k$$

Proof. Q.E.D.

□

References

.

Angrist, J. D. and G. W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Eects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431-442.

Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002, June). Models and issues in data stream systems. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 1-16). ACM.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012), "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica* 80, 2369-2429.

Belloni, A., and V. Chernozhukov (2009): "Computational Complexity of MCMC-based Estimators in Large Samples," *Annals of Statistics* 37, 2011-2055.

Belloni, A., C. Hansen, and I. Fernandez-Val (2013a). "Program Evaluation with High-Dimensional Data," MIT Working Paper.

Belloni, A., V. Chernozhukov. and C. Hansen (2013b): "Inference on Treatment Effects After Selection Amongst High-Dimensional Controls (with an Application to Abortion and Crime)," *Review of Economic Studies*, forthcoming.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004): "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119, 249-275.

Brown, R. G. and Meyer, R. F. 1961, "The fundamental theorem of exponential smoothing," *Operations Research*, 9(5), 673-685.

Couper, M. P. (2005): "Technology Trends in Survey Data Collection," *Social Science Computer Review*, 23, 486-501.

Dragoman, D. and M. Dragoman (2004): *Quantum-Classical Analogies* (Springer, Berlin).

Einav, L. and J. Levin (2014): "The Data Revolution and Economic Analysis," in *Innovation Policy and the Economy*, Volume 14, ed. by J. Lerner and S. Stern, forthcoming.

Favre, Maroussia, Amrei Wittwer, Hans Rudolf Heinimann, Vyacheslav I. Yukalov, and Didier Sornette (2016). "Quantum Decision Theory in Simple Risky Choices," *PLOS One*, https://doi.org/10.1371/journal.pone.0168045

Fowler, J. H. and N. A. Christakis (2008): "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis over 20 years in the Framingham Heart Study," *BMJ: British Medical Journal*, 337, 1-9.

Green, J. Hilary, and I. A. Macdonald (1981). "The influence of intravenous glucose on body temperature." *Quarterly Journal of Experimental Physiology* 66(4), 465-473.

Hacohen-Gourgy, S., L. S. Martin, E. Flurin, V. V. Ramasesh, K. B. Whaley and I. Siddiqi, "Dynamics of simultaneously measured non-commuting observables," *Nature,* doi:10.1038/nature19762 (October 2016).

Hansen, Ben B. and Jake Bowers (2008). "Covariate Balance in Simple, Stratified and Clustered Comparative Studies." *Statistical Science*, 23, 219–236.

Holland, P. W. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945-960.

Hui, Xiaonan, Jinhai Zhou, Chen Xu, Shilie Zheng, Hao Chi, Xiaofeng Jin, and Xianmin Zhang (2013). "A real-time detection and self-control phase-sensitive OTDR distributed sensor system." In *12th International Conference on Optical Communications and Networks (ICOCN).*

Imbens, Guido W (2000). "The role of the propensity score in estimating dose-response functions." *Biometrika*, 87(3), 706-710.

Imbens, Guido W., and Donald B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press, New York.

Imbens, Guido W., and Jeffrey M. Wooldridge (2009). "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature* 47(1), 5-86.

Jin, H. and D. B. Rubin (2008): "Principal Stratication for Causal Inference with Extended Partial Compliance," *Journal of the American Statistical Association*, 103, 101-111.

Kamenica, E., S. Mullainathan and R. H. Thaler. (2011): "Helping Consumers Know Themselves," *American Economic Review: Papers & Proceedings* 101, 417-422.

King, G. (2011): "Ensuring the Data Rich Future of the Social Sciences, *Science,* 331, 719-721.

Morgenstern, Oskar. (1950). "On the accuracy of economic observations: Foreign trade statistics", in: *The Accuracy of Economic Observations*, Chapter IX, Princeton University Press.

Nadungodage, Chandima Hewa, Yuni Xia, Fang Li, Jaehwan John Lee, and Jiaqi Ge (2011). "StreamFitter: a real time linear regression analysis system for continuous data streams." In *International Conference on Database Systems for Advanced Applications*, pp. 458-461. Springer Berlin Heidelberg, 2011.

Neyman, J. 1923 [1990]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science* 5, 465–472.

Orphanides, Athanasios (2001). "Monetary Policy Rules Based on Real-Time Data," *American Economic Review* 91(4), 964-985.

Rubin, D. (1974): "Estimating Causal Impacts of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66,688-701.

Stern A. D., B. A. Alexander, and A, Chandra (2017). "The Economics of Precision Medicine." *Science*, forthcoming.

Suits, D. B. (1957), "Use of Dummy Variables in Regression Equations," *Journal of the American Statistical Association*, 52, 548-551.

Suits, D. B. (1984), "Dummy Variables: Mechanics v. Interpretation," *Review of Economics and Statistics*, 66, 177-180.

Varian, H. R. (2014): "Big Data: New Tricks for Ecoometrics," *Journal of Economic Perspectives*, 28(2): 3-28.

Van der Wal, C. H., A. C. J. ter Haar, F. K. Wilhelm, R. N. Schouten, C. J. P. M. Harmans, T. P. Orlando, S. Lloyd, and J. E. Mooij, (2000): "Quantum Superposition of Macroscopic Persistent-Current States," *Science* 290, 773-777.

Van Zandt, Timothy (1999). "Real-time decentralized information processing as a model of organizations with boundedly rational agents." *The Review of Economic Studies* 66(3), 633-658.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data.* MIT press.

Yukalov V. I., and D. Sornette (2014). Manipulating decision making of typical agents. *IEEE Transactions on Systems, Man and Cybernetics: Systems.* 44:1155–1168.

Zhu, Yunyue, and Dennis Shasha. "Statstream: Statistical monitoring of thousands of data streams in real time." *Proceedings of the 28th international conference on Very Large Data Bases.* VLDB Endowment, 2002.

Zwane, A. P., J. Zinman, E. Van Dusene, W. Paiente, C. Null, E. Miguel, M. Kremer, D. S. Karlan, R. Hornbeck, X. Giné, E. Duflo, F. Devoto, B. Crepon, and A. Banerjee, (2011): "Being Surveyed Can Change Later Behavior and Related Parameter Estimates," *Proceedings of the National Academy of Sciences,* 108, 1821–1826.