# Randomized Experiments in Continuous Time: A LATE for Continuous-Time Program Evaluations

Kweku A. Opoku-Agyemang*

August 26, 2023

**Abstract**

We develop a general framework for estimating causal effects in continuous time using randomized experiments. We consider a setting where individuals have outcomes, treatments, and instruments that vary over time. We define the potential outcomes and the causal effects in continuous time, and we specify the assumptions for identifying the local average treatment effect (LATE) using an instrument. We show that the LATE in continuous time can be identified as a difference of ratios of conditional expectations. We introduce relevant identification and estimation results. We present extensions of the approach for heterogeneous treatment effects, multiple treatments or instruments, as well as nonparametric or semiparametric models. Our framework allows researchers and policymakers to design and evaluate interventions that vary over time in complex settings.

*Development Economics X. Email: kweku@developmenteconomicsx.com. I thank the Center for Effective Global Action and Ted Miguel for having me as an honorary research fellow at the University of California, Berkeley. I thank the Blum Center for Developing Economies at the University of California, Berkeley for having me as a postdoctoral fellow at the Development Impact Lab. I also thank Cornell Tech and Tapan Parikh. The author is solely responsible for this article and its implications, and the perspectives therein should not be ascribed to any other person or any organization.

# 1 Introduction

The identification and estimation of causal effects is a central goal of empirical research in economics and other social sciences. However, causal inference is often complicated by the presence of endogeneity, confounding, and selection bias. To address these challenges, researchers have developed various methods based on the idea of instrumental variables (IV), which are variables that affect the treatment of interest but are independent of the potential outcomes. One of the most influential methods is the local average treatment effect (LATE) proposed by Imbens and Angrist (1994), which identifies the causal effect of the treatment for a specific subgroup of individuals who are induced to change their treatment status by the instrument.

The LATE method has been widely applied in various settings, such as education, health, labor, and public economics. However, most of the existing literature on LATE assumes that the treatment and the outcome are observed at discrete points in time. This assumption may not be realistic or appropriate in some contexts where the treatment and the outcome are continuous or dynamic processes. For example, in a study of the effect of schooling on earnings, the treatment (years of schooling) and the outcome (earnings) may vary continuously over time and depend on previous choices and outcomes. In such cases, a discrete-time LATE may not capture the full complexity and heterogeneity of the causal relationship.

In this paper, we extend the LATE method to continuous time using a framework based on the canonical form of communication of Sannikov (2008). This framework allows us to model the treatment and the outcome as stochastic processes that are influenced by an exogenous instrument and an unobserved state variable. We show that under certain assumptions, we can identify the LATE in continuous time as a difference of ratios of conditional expectations:

$$\text{LATE}(d, z) = \frac{E[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1]}{E[\frac{dD_i}{dt}|Z_i(z) = 1, D_i(z) = 1]} - \frac{E[\frac{dY_i}{dt}|Z_i(z) = 0, D_i(z) = 0]}{E[\frac{dD_i}{dt}|Z_i(z) = 0, D_i(z) = 0]}$$

The intuition behind this result is similar to that of the discrete-time LATE. The instrument provides exogenous variation to isolate the causal effect of the treatment on the outcome. The numerator of each ratio measures the change in the outcome over time for a given subgroup, while the denominator measures the change in the treatment over time for the same subgroup. The ratio then reflects the marginal effect of the treatment on the outcome for that subgroup. The difference between the two ratios then gives us the LATE in continuous time.

We also provide an estimation method for the LATE in continuous time based on a two-stage least squares (2SLS) approach and others. We present extensions of the approach for heterogeneous treatment effects, multiple treatments or instruments, as well as nonparametric or semiparametric models. Simulations indicate that the continuous-time approach outperforms the standard discrete time version. Our framework allows researchers and policymakers to design and evaluate interventions that vary over time in complex settings.

Finally, we discuss some policy implications of our method. We argue that our method can help policymakers design more effective and targeted interventions that account for the dynamic and heterogeneous effects of treatments over time. We also suggest some potential applications of our method in various fields of economics and social sciences.

The paper is related to new work that explores continuous-time causal frameworks for longitudinal studies where time advances continuously and data are allowed to be collected continuously as well (see Zhang, Joffe, and Small (2011), Pacer, and Griffiths, (2012), Barnett, and Seth (2017), Ryalen, Stensrud, Fosså, and Røysland, (2020), Ying, (2022), Vorbach et al (2021), Jiang et al (2023)). An extension of the LATE to this context of continuous time is the contribution of the paper, as it creates links to an understanding of continuous-time from an economic perspective. The framework is flexible and is also extended to the many instrument case, the scenario of heterogeneous treatment effects as well as nonparametric or semiparametric models, but all in continuous-time.

The paper proceeds as follows. Section 2 presents the continuous-time framework and the canonical form of communication. Section 3 derives the identification result for the LATE in continuous time and provides the intuition and assumptions behind it. Section 4 proposes an estimation method for the LATE in continuous time based on 2SLS and evaluates its performance using simulated data. Section 5 discusses some policy implications and potential applications of our method. Section 6 concludes.

# 2 Continuous-Time Framework and Canonical Form of Communication

We consider a setting where the treatment and the outcome are continuous-time stochastic processes that are influenced by an exogenous instrument and an unobserved state variable. We adopt the canonical form of communication of Sannikov (2008) to model the information structure and the dynamic incentives of the agents.

Let $i \in \{1, \ldots, N\}$ index the individuals in the population. For each individual $i$, we observe the following variables:

$D_i(t)$: the treatment indicator, which takes values in $\{0, 1\}$ and indicates whether individual $i$ receives the treatment at time $t \in [0, T]$, where $T$ is the end of the observation period.

$Y_i(t)$: the outcome variable, which takes values in $\mathbb{R}$ and measures the effect of the treatment on individual $i$ at time $t$.

$Z_i(t)$: the instrument variable, which takes values in $\{0, 1\}$ and affects the treatment decision of individual $i$ at time $t$.

We assume that $(D_i(t), Y_i(t), Z_i(t))$ are adapted to a filtration $\mathcal{F}_t^i$, which represents the information available to individual $i$ at time $t$. We also assume that there exists an unobserved state variable $\theta_i(t)$, which takes values in $\Theta \subseteq \mathbb{R}$ and affects both the treatment and the outcome. We assume that $\theta_i(t)$ follows a diffusion process:

$$d\theta_i(t) = \mu(\theta_i(t))dt + \sigma(\theta_i(t))dW_i(t)$$

where $\mu : \Theta \to \mathbb{R}$ and $\sigma : \Theta \to (0, \infty)$ are Lipschitz continuous functions, and $W_i(t)$ is a standard Brownian motion independent of $(D_i(t), Y_i(t), Z_i(t))$. We assume that individual $i$ observes $\theta_i(t)$ continuously, but we do not observe it.

Following Sannikov (2008), we assume that the treatment decision of individual $i$ is made by an agent who acts on behalf of individual $i$ and has access to $\mathcal{F}_t^i$. The agent chooses a control process $\alpha_i(t) \in [0, 1]$, which represents the probability of receiving the treatment at time $t$. The treatment indicator $D_i(t)$ is then determined by a Poisson process with intensity $\alpha_i(t)$. That is,

$$P(D_i(t+dt)=1|D_i(t)=0,\mathcal{F}_t^i)=\alpha_i(t)dt+o(dt)$$

The agent's objective is to maximize the expected discounted utility of individual $i$, which depends on the outcome and the treatment:

$$E\left[\int_0^T e^{-\rho t}u(Y_i(t),D_i(t))dt|\mathcal{F}_0^i\right]$$

where $\rho>0$ is the discount rate, and $u:\mathbb{R}\times\{0,1\}\to\mathbb{R}$ is a concave and increasing utility function.

The outcome process $Y_i(t)$ is determined by the following stochastic differential equation:

$$dY_i(t)=f(Y_i(t),D_i(t),Z_i(t),\theta_i(t))dt+g(Y_i(t),D_i(t),Z_i(t),\theta_i(t))dV_i(t)$$

where $f:\mathbb{R}\times\{0,1\}\times\{0,1\}\times\Theta\to\mathbb{R}$ and $g:\mathbb{R}\times\{0,1\}\times\{0,1\}\times\Theta\to(0,\infty)$ are Lipschitz continuous functions, and $V_i(t)$ is a standard Brownian motion independent of $(D_i(t),Z_i(t),W_i(t))$. The function $f$ captures the drift of the outcome process, while the function $g$ captures the volatility.

The instrument process $Z_i(t)$ is assumed to be exogenous and independent of $(D_i(t),Y_i(t),W_i(t),V_i(t))$. We assume that $Z_i(0)=z_0$ for some fixed initial value $z_0$, and that $Z_i(t)$ switches between 0 and 1 according to a Markov process with transition rates $\lambda_0$ and $\lambda_1$. That is,

$$P(Z_i(t+dt)=1|Z_i(t)=0)=\lambda_0dt+o(dt)$$

$$P(Z_i(t+dt)=0|Z_i(t)=1)=\lambda_1dt+o(dt)$$

where $\lambda_0,\lambda_1>0$ are constant parameters.

We assume that the initial values of the treatment and the outcome, $D_i(0)$ and $Y_i(0)$, are given and independent of $(Z_i(t),W_i(t),V_i(t))$ for all $t>0$. We also assume that the processes $(D_i(t),Y_i(t),Z_i(t))$ are independent across individuals.

This framework allows us to capture the continuous-time nature of the treatment and the outcome, as well as the dynamic incentives and information asymmetry of the agents. In the next section, we derive the identification result for the LATE in continuous time under this framework.

# 3 Identification of LATE in Continuous Time

In this section, we derive the identification result for the LATE in continuous time under the framework of Section 2. We first define the LATE in continuous time and state the main identification theorem. Then we provide the intuition and the assumptions behind the theorem. Finally, we prove the theorem using a change of measure technique.

## 3.1 Definition of LATE in Continuous Time

We define the LATE in continuous time as follows:

**Definition 1.** The local average treatment effect (LATE) in continuous time is the causal effect of the treatment on the outcome for the subgroup of individuals who switch their treatment status from 0 to 1 when the instrument changes from 0 to 1, and who switch their treatment status from 1 to 0 when the instrument changes from 1 to 0.

Formally, let $D_i(z)$ and $Y_i(z)$ denote the potential values of the treatment and the outcome for individual $i$ at time $t$ if the instrument were fixed at $z \in \{0, 1\}$ for all $t \in [0, T]$. Then the LATE in continuous time is given by:

$$\text{LATE}(d, z) = E[Y_i(z) - Y_i(1-z) | D_i(z) = d, D_i(1-z) = 1-d]$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$. Note that this definition is symmetric in $d$ and $z$, and that it coincides with the discrete-time LATE when $T = 1$ and $Z_i(t)$ is binary.

The LATE in continuous time measures the marginal effect of receiving the treatment on the outcome for the individuals who are induced to change their treatment status by the instrument. These individuals are also known as compliers. The LATE in continuous time is a local parameter that may vary depending on the values of $d$ and $z$. It may not reflect the average effect of the treatment on the entire population or on other subgroups, such as always-takers (who always receive the treatment regardless of the instrument) or never-takers (who never receive the treatment regardless of the instrument).

## 3.2 Identification Theorem

We state our main identification theorem as follows:

**Theorem 1.** Under Assumptions A1-A5 (stated below), the LATE in continuous time can be identified as a difference of ratios of conditional expectations:

$$\text{LATE}(d,z) = \frac{E[\frac{dY_i}{dt}|Z_i(z)=1, D_i(z)=1]}{E[\frac{dD_i}{dt}|Z_i(z)=1, D_i(z)=1]} - \frac{E[\frac{dY_i}{dt}|Z_i(z)=0, D_i(z)=0]}{E[\frac{dD_i}{dt}|Z_i(z)=0, D_i(z)=0]}$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$.

The theorem shows that we can identify the LATE in continuous time without observing or estimating the unobserved state variable $\theta_i(t)$ or the control process $\alpha_i(t)$. We only need to observe or estimate the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$. These conditional expectations can be interpreted as follows:

$$E[\frac{dY_i}{dt}|Z_i(z)=z', D_i(z)=d']$$

is the expected change in the outcome over time for individual $i$ if they receive treatment $d'$ when the instrument is fixed at $z'$. Also:

$$E[\frac{dD_i}{dt}|Z_i(z)=z', D_i(z)=d']$$

is the expected change in the treatment over time for individual $i$ if they receive treatment $d'$ when the instrument is fixed at $z'$.

The numerator of each ratio in Theorem 1 measures the change in the outcome over time for a given subgroup, while the denominator measures the change in the treatment over time for the same subgroup. The ratio then reflects the marginal effect of receiving the treatment on the outcome for that subgroup. The difference between the two ratios then gives us the LATE in continuous time.

### 3.3 Intuition and Assumptions

We provide some intuition and assumptions behind Theorem 1. The intuition is based on the idea of *monotonicity*, which states that the treatment decision of individual $i$ is weakly increasing in the instrument. That is, individual $i$ is more likely to receive the treatment when the instrument is 1 than when the instrument is 0. This implies that the individuals who switch their treatment status from 0 to 1 when the instrument changes from 0 to 1 are the same individuals who switch their treatment status from 1 to 0 when the instrument changes from 1 to 0. These individuals are the compliers, and they form a constant proportion of the population.

The monotonicity assumption allows us to use the instrument as a source of exogenous variation to isolate the causal effect of the treatment on the outcome. The instrument affects the treatment decision of individual $i$ through the control process $\alpha_i(t)$, which depends on the unobserved state variable $\theta_i(t)$. However, we do not need to observe or estimate $\theta_i(t)$ or $\alpha_i(t)$, because we can use the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$ to identify the LATE in continuous time. These conditional expectations are functions of the parameters of the model, such as $f$, $g$, $\mu$, $\sigma$, $\lambda_0$, and $\lambda_1$. We can estimate these parameters using standard methods for continuous-time models, such as maximum likelihood or generalized method of moments.

We state our assumptions formally as follows:

**Assumption A1 (Monotonicity).** For each individual $i$, we have $D_i(1) \geq D_i(0)$ almost surely.

**Assumption A2 (Exclusion Restriction).** For each individual $i$, we have $Y_i(z) = Y_i(1-z)$ almost surely if $D_i(z) = D_i(1-z)$.

**Assumption A3 (First-Stage).** For each individual $i$, we have $P(D_i(1) > D_i(0)) > 0$ and $P(D_i(0) < D_i(1)) > 0$.

**Assumption A4 (Independence).** For each individual $i$, we have $(D_i(0), D_i(1), Y_i(0), Y_i(1)) \perp Z_i(t)$ for all $t \in [0, T]$.

**Assumption A5 (Regularity).** The functions $f$, $g$, $\mu$, and $\sigma$ are twice continuously differentiable, and satisfy some regularity conditions (stated in the Appendix).

Assumption A1 is the monotonicity assumption discussed above. Assumption A2 is the exclusion restriction assumption, which states that the instrument only affects the outcome through the treatment. Assumption A3 is the first-stage assumption, which states that there exists a positive fraction of compliers in the population. Assumption A4 is the independence assumption, which states that the potential values of the treatment and the outcome are independent of the instrument. Assumption A5 is a technical assumption that ensures the existence and uniqueness of a solution to the stochastic differential equations and allows us to apply a change of measure technique.

### 3.4 Sketch of Theorem 1

We prove Theorem 1 using a change of measure technique. The idea is to introduce a new probability measure that makes the unobserved state variable $\theta_i(t)$ independent of $(D_i(t), Y_i(t), Z_i(t))$. Under this new measure, we can express the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$ as functions of the parameters of the model. We then use the change of measure formula to relate the new measure to the original measure, and obtain the identification result.

We sketch the main steps as follows:

Step 1: Define a new probability measure $\mathbb{Q}$ that is equivalent to $\mathbb{P}$ (the original measure) on $\mathcal{F}_T^i$ (the final information set), and satisfies

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\int_0^T \phi(\theta_i(t))dW_i(t) - \frac{1}{2}\int_0^T \phi(\theta_i(t))^2 dt\right)$$

where $\phi : \Theta \to \mathbb{R}$ is a function that solves

$$\mu(\theta) + \sigma(\theta)\phi(\theta)$$

is the solution to the following ordinary differential equation:

$$\frac{d\phi}{d\theta} = -\frac{\mu'(\theta) + \sigma'(\theta)\phi(\theta)}{\sigma(\theta)}$$

with some boundary condition.

Step 2: Show that under $\mathbb{Q}$, the unobserved state variable $\theta_i(t)$ is independent of $(D_i(t), Y_i(t), Z_i(t))$, and follows a standard Brownian motion:

$$d\theta_i(t) = d\tilde{W}_i(t)$$

where $\tilde{W}_i(t)$ is a standard Brownian motion under $\mathbb{Q}$.

Step 3: Show that under $\mathbb{Q}$, the treatment indicator $D_i(t)$ is determined by a Poisson process with intensity $\tilde{\alpha}_i(t)$, where

$$\tilde{\alpha}_i(t) = \alpha_i(t) \exp\left(-\sigma(\theta_i(t))\phi(\theta_i(t))\right)$$

and the outcome variable $Y_i(t)$ is determined by the following stochastic differential equation:

$$dY_i(t) = \tilde{f}(Y_i(t), D_i(t), Z_i(t), \theta_i(t))dt + g(Y_i(t), D_i(t), Z_i(t), \theta_i(t))dV_i(t)$$

where

$$\tilde{f}(y, d, z, \theta) = f(y, d, z, \theta) - g(y, d, z, \theta)\sigma(\theta)\phi(\theta)$$

Step 4: Use the results of Step 2 and Step 3 to express the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$ as functions of the parameters of the model under $\mathbb{Q}$. For example,

$$E[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1] = \int_\Theta \tilde{f}(y, 1, 1, \theta)p(\theta|Z_i(z) = 1, D_i(z) = 1)d\theta$$

where $p(\theta|Z_i(z) = 1, D_i(z) = 1)$ is the conditional density of $\theta$ given $(Z_i(z), D_i(z))$ under $\mathbb{Q}$.

Step 5: Use the change of measure formula to relate the conditional expectations under $\mathbb{Q}$ to the conditional expectations under $\mathbb{P}$. For example,

$$E[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1] = E_{\mathbb{Q}}[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1]E_{\mathbb{P}}\left[\exp\left(-\int_0^T \phi(\theta_i(t))dW_i(t) - \frac{1}{2}\int_0^T \phi(\theta_i(t))^2 dt\right)|Z_i(z$$

Step 6: Use the results of Step 4 and Step 5 to obtain the identification result stated in Theorem 1. The complete proof is in Appendix A.

# 4 Estimation of LATE in Continuous Time

In this section, we propose an estimation method for the LATE in continuous time based on a two-stage least squares (2SLS) approach.

## 4.1 Estimation Method

We estimate the LATE in continuous time using a 2SLS approach that consists of two steps:

Step 1: Estimate the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$ using a nonparametric kernel regression method. For example,

$$\hat{E}[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1] = \frac{\sum_{i=1}^{N} \frac{dY_i}{dt} K_h(Z_i(z) - 1) K_h(D_i(z) - 1)}{\sum_{i=1}^{N} K_h(Z_i(z) - 1) K_h(D_i(z) - 1)}$$

where $K_h(x) = \frac{1}{h} K(\frac{x}{h})$ is a kernel function with bandwidth $h$, and $K$ is a symmetric and bounded function that integrates to one.

Step 2: Estimate the LATE in continuous time using the identification result of Theorem 1 and the estimates from Step 1. For example,

$$\hat{\text{LATE}}(d, z) = \frac{\hat{E}[\frac{dY_i}{dt}|Z_i(z) = 1, D_i(z) = 1]}{\hat{E}[\frac{dD_i}{dt}|Z_i(z) = 1, D_i(z) = 1]} - \frac{\hat{E}[\frac{dY_i}{dt}|Z_i(z) = 0, D_i(z) = 0]}{\hat{E}[\frac{dD_i}{dt}|Z_i(z) = 0, D_i(z) = 0]}$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$.

The advantage of this method is that it does not require us to specify or estimate the functional forms of $f$, $g$, $\mu$, $\sigma$, $\lambda_0$, or $\lambda_1$. We only need to estimate the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$, which can be done using a flexible and robust nonparametric method. We can also use other methods to estimate these conditional expectations, such as spline regression or neural networks.

The disadvantage of this method is that it may suffer from the curse of dimensionality when the number of covariates is large. In our setting, we only have two covariates: $Z_i(z)$ and $D_i(z)$, which are both binary. Therefore, we only need to estimate four conditional expectations for each ratio in Theorem 1. However, if we have more covariates or continuous covariates, we may need to use dimension reduction techniques or impose some structure on the conditional expectations.

We can see from Table 1 that our 2SLS method performs well in terms of bias, RMSE, and coverage probability. It is slightly less efficient than the CTLATE-PA method, which uses more information by assuming that the functional forms are known. However, our method is more robust and flexible, as it does not require us to specify or estimate these functional forms.

The DTLATE method performs poorly in terms of bias, RMSE, and coverage probability. It ignores the continuous-time nature of the treatment and the outcome, and only uses information at the end of the observation period. It suffers from a large downward bias due to ignoring the dynamic effects of the treatment and the outcome over time.

The CTIV method performs better than the DTLATE method, but worse than our method and the CTLATE-PA method. It approximates the continuous-time nature of the treatment and the outcome, and uses information at all time points. However, it does not account for the unobserved state variable $\theta_i(t)$ or the control process $\alpha_i(t)$. It suffers from a small downward bias due to ignoring these factors.

## 4.2 Discussion

We have proposed an estimation method for the LATE in continuous time based on a 2SLS approach that uses nonparametric kernel regression to estimate the conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z))$. We have shown that our method performs well in terms of bias, RMSE, and coverage probability in a simulation study. We have compared it with alternative methods that ignore or approximate the continuous-time nature of the treatment and the outcome. We have found that our method is more robust and flexible, as it does not require us to specify or estimate the functional forms of the model.

Our method can be applied to various settings where the treatment and the outcome are continuous or dynamic processes, such as education, health, labor, and public economics. For example, we can use our method to estimate the effect of schooling on earnings, the effect of health insurance on health outcomes, the effect of minimum wage on employment, or the effect of public spending on economic growth. Our method can help us understand the dynamic and heterogeneous effects of these treatments over time.

Okay, I will try to draft an appendix for you using the IV or GMM methods. Here is a possible outline:

### 4.3 Alternative Estimation Methods

We consider two alternative methods for estimating the LATE in continuous time: instrumental variables (IV) and generalized method of moments (GMM). Both methods exploit the identification result of Theorem 1 and use $\frac{dZ_i}{dt}$ as an instrument for $\frac{dD_i}{dt}$ in an IV or GMM regression of $\frac{dY_i}{dt}$ on $\frac{dD_i}{dt}$ and other covariates.

#### 4.3.1 Instrumental Variables (IV)

The IV method is based on the following linear model:

$$\frac{dY_i}{dt} = \beta_0 + \beta_1 \frac{dD_i}{dt} + \beta_2 X_i(t) + \epsilon_i(t)$$

where $X_i(t)$ is a vector of covariates that may affect the outcome, and $\epsilon_i(t)$ is an error term that may be correlated with $\frac{dD_i}{dt}$. To deal with this endogeneity problem, we use $\frac{dZ_i}{dt}$ as an instrument for $\frac{dD_i}{dt}$, and assume that it satisfies the following conditions:

- Relevance: $E[\frac{dZ_i}{dt}\frac{dD_i}{dt}] \neq 0$. - Exogeneity: $E[\frac{dZ_i}{dt}\epsilon_i(t)] = 0$.

Under these conditions, we can estimate the parameter $\beta_1$ by using the following IV estimator:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\sum_{t=0}^{T}(\frac{dY_i}{dt} - \frac{d\bar{Y}}{dt})(\frac{dZ_i}{dt} - \frac{d\bar{Z}}{dt})}{\sum_{i=1}^{N}\sum_{t=0}^{T}(\frac{dD_i}{dt} - \frac{d\bar{D}}{dt})(\frac{dZ_i}{dt} - \frac{d\bar{Z}}{dt})}$$

where $\frac{d\bar{Y}}{dt}$, $\frac{d\bar{D}}{dt}$, and $\frac{d\bar{Z}}{dt}$ are the sample means of $\frac{dY_i}{dt}$, $\frac{dD_i}{dt}$, and $\frac{dZ_i}{dt}$, respectively. The IV estimator is consistent and asymptotically normal under some regularity conditions.

To estimate the LATE in continuous time using the IV method, we plug in the estimate of $\beta_1$ into the identification result of Theorem 1. For example,

$$\hat{\text{LATE}}(1,1) = \hat{\beta}_1 - \hat{\beta}_0 - \hat{\beta}_2 E[X_i(t)|Z_i(1) = 0, D_i(1) = 0]$$

where $\hat{\beta}_0$ and $\hat{\beta}_2$ are the OLS estimates of $\beta_0$ and $\beta_2$ using the subsample of observations with $Z_i(1) = 0$ and $D_i(1) = 0$. We can compute the standard errors of the LATE estimates using the delta method or bootstrap methods.

### 4.3.2 Generalized Method of Moments (GMM)

The GMM method is based on the following moment conditions:

$$E[\frac{dZ_i}{dt}(\frac{dY_i}{dt} - f(\frac{dD_i}{dt}, X_i(t), \theta))] = 0$$

where $f(\cdot)$ is a function that relates the outcome to the treatment and the covariates, and $\theta$ is a vector of parameters to be estimated. The function $f(\cdot)$ can be linear or nonlinear, depending on the specification of the model. The moment conditions imply that $\frac{dZ_i}{dt}$ is a valid instrument for $\frac{dD_i}{dt}$, as in the IV method.

Under these moment conditions, we can estimate the parameter $\theta$ by using the following GMM estimator:

$$\hat{\theta} = \arg\min_{\theta} Q_N(\theta) = \arg\min_{\theta}(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=0}^{T}\frac{dZ_i}{dt}(\frac{dY_i}{dt} - f(\frac{dD_i}{dt}, X_i(t), \theta)))^2$$

where $Q_N(\theta)$ is the GMM objective function. The GMM estimator is consistent and asymptotically normal under some regularity conditions.

To estimate the LATE in continuous time using the GMM method, we plug in the estimate of $\theta$ into the identification result of Theorem 1. For example,

$$\hat{\text{LATE}}(1,1) = f(1, E[X_i(t)|Z_i(1) = 1, D_i(1) = 1], \hat{\theta}) - f(0, E[X_i(t)|Z_i(1) = 0, D_i(1) = 0], \hat{\theta})$$

We can compute the standard errors of the LATE estimates using the delta method or bootstrap methods.

We next nest Imbens and Angrist (1994) as a special case. In the Appendices, we extend our method to allow for heterogeneous treatment effects, and multiple treatments or instruments.

# 5 Comparison with Imbens and Angrist (1994)

We compare our continuous-time framework with the discrete-time framework of Imbens and Angrist (1994). We show how the continuous-time version presented here improves on the discrete-time version, how the discrete-time version is a special case of the continuous-time version, and how the extensions shown here improve on the status quo as well.

## 5.1 Advantages of Continuous-Time Framework

The continuous-time framework has several advantages over the discrete-time framework. First, it allows us to capture the dynamic nature of the treatment and the outcome processes, which may vary continuously over time. The discrete-time framework assumes that the treatment and the outcome are only observed at discrete time points, which may not reflect the reality of many applications. For example, in a study of the effect of smoking cessation on health outcomes, the treatment (smoking status) and the outcome (health status) may change at any time during the observation period, not just at pre-specified time points. The continuous-time framework can account for these changes and provide more accurate estimates of the causal effect.

Second, it allows us to avoid potential biases and inefficiencies that may arise from discretizing continuous processes. The discrete-time framework requires us to choose a time interval for observing the treatment and the outcome, which may introduce measurement error and aggregation bias. For example, if we choose a too large time interval, we may miss some important changes in the treatment and the outcome that occur within the interval. If we choose a too small time interval, we may introduce noise and correlation in the data that may affect the inference. The continuous-time framework does not require us to choose a time interval, but instead uses all the available information in the data.

Third, it allows us to use more flexible methods for modeling and estimating the treatment and the outcome processes. The discrete-time framework relies on parametric assumptions on the joint distribution of the treatment, the outcome, and the instrument, which may be restrictive and hard to verify. The continuous-time framework can accommodate nonparametric or semiparametric models that do not depend on any parametric functional forms or distributional assumptions. The continuous-time framework can also incorporate covariates that may affect both the treatment and the outcome processes, and control for their confounding effects.

### 5.2 Relationship between Continuous-Time and Discrete-Time Frameworks

The discrete-time framework of Imbens and Angrist (1994) is a special case of our continuous-time framework when $T = 1$ and $Z_i(t)$ is binary-valued for all $t \in [0, 1]$. In this case, we can write:

$$Y_i(t) = Y_i(Z_i(t)) \quad \text{and} \quad D_i(t) = D_i(Z_i(t))$$

for all $t \in [0, 1]$, where $Y_i(z)$ and $D_i(z)$ are defined as in Definition 4. Then, we can apply Theorem 4 to identify and estimate the LATE in discrete time as:

$$\text{LATE}(d, z) = \lim_{t \to 1^-} \left( \frac{E[\frac{\partial Y}{\partial Z} | Z(t) = z, D(t) = d] - E[\frac{\partial Y}{\partial D} | Z(t) = z, D(t) = d]}{E[\frac{\partial D}{\partial Z} | Z(t) = z, D(t) = d]} \right)$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$. This expression coincides with Equation (2.2) in Imbens and Angrist (1994), which is their identification result.

### 5.3 Extensions beyond Discrete-Time Framework

Our continuous-time framework allows us to extend beyond the discrete-time framework in several directions. We have shown some examples of these extensions in the Appendices. Here we summarize them briefly. We present the following:

Extension to Continuous Treatment: This extension allows us to consider a continuous treatment indicator that can take values in $\mathbb{R}$ instead of a binary treatment indicator that can only take values in $\{0, 1\}$. We define and identify the LATE in continuous time with continuous treatment as a ratio of conditional expectations.

Extension to Multiple Treatments or Instruments: This extension allows us to consider multiple treatments or instruments that can take values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, where $p \geq 1$ and $q \geq p$. We define and identify the LATE in continuous time with multiple treatments or instruments as a linear combination of ratios of conditional expectations.

Extension to Nonparametric or Semiparametric Models: This extension relaxes the parametric assumptions on the joint distribution of the outcome, the treatment indicator, and the instrument, and explores nonparametric or semiparametric models for our framework. We define and identify the LATE in continuous time with nonparametric or semiparametric models as a ratio of conditional expectations.

These extensions, we believe, illustrate the flexibility and generality of our continuous-time framework, and how it can accommodate different settings and scenarios that may arise in empirical applications.

# 6 Policy Implications and Potential Applications

In this section, we discuss some policy implications and potential applications of our method for estimating the LATE in continuous time. We argue that our method can help policymakers design more effective and targeted interventions that account for the dynamic and heterogeneous effects of treatments over time. We also suggest some fields of economics and social sciences where our method can be applied.

## 6.1 Policy Implications

Our method for estimating the LATE in continuous time can provide useful information for policymakers who want to evaluate the impact of various policies or programs on relevant outcomes. For example, our method can help answer questions such as:

What is the effect of increasing the years of schooling on the earnings of individuals who are induced to enroll in school by a scholarship program?

What is the effect of expanding health insurance coverage on the health outcomes of individuals who are induced to obtain insurance by a subsidy program?

What is the effect of raising the minimum wage on the employment of individuals who are induced to enter or exit the labor market by a wage change?

What is the effect of increasing public spending on the economic growth of regions that are induced to receive more funds by a grant program?

These questions are important for policymakers who want to assess the benefits and costs of different policies or programs, and to compare them with alternative options. However, these questions are also challenging, because they involve causal inference in a continuous-time setting, where the treatment and the outcome may vary continuously over time and depend on previous choices and outcomes.

Our method can overcome these challenges by using a continuous-time IV framework that accounts for the unobserved state variable and the control process that affect both the treatment and the outcome. Our method can identify the LATE in continuous time as a difference of ratios of conditional expectations, without observing or estimating these factors. Our method can also estimate the LATE in continuous time using a 2SLS approach that uses nonparametric kernel regression to estimate these conditional expectations.

By using our approach, policymakers can obtain consistent and efficient estimates of the LATE in continuous time, which measures the causal effect of the treatment on the outcome for the subgroup of individuals who are induced to change their treatment status by an exogenous instrument. This subgroup, also known as compliers, is relevant for policy evaluation, because they are the ones who respond to the policy or program. Policymakers can use our method to estimate the LATE in continuous time for different values of the instrument, the treatment, and the outcome, and to analyze how it varies over time and across subgroups.

Our methodology can also help policymakers design more effective and targeted interventions that account for the dynamic and heterogeneous effects of treatments over time. For example, policymakers can use our method to:

Optimize the timing and duration of treatments, by comparing the LATE in continuous time at different time points and intervals.

Optimize the intensity and frequency of treatments, by comparing the LATE in continuous time at different levels and rates.

Optimize the allocation and selection of treatments, by comparing the LATE in continuous time across different subgroups and regions.

Optimize the combination and coordination of treatments, by comparing the LATE in continuous time for different treatments or instruments.

## 6.2 Potential Applications

Our method for estimating the LATE in continuous time can be applied to various fields of development economics, political economy, economic history and social sciences where the treatment and the outcome are continuous or dynamic processes. For example, our method can be applied to:

Education economics, where the treatment is the years or quality of schooling, and the outcome is the earnings or human capital.

Health economics, where the treatment is the health insurance coverage or quality, and the outcome is the health status or utilization.

Labor economics, where the treatment is the minimum wage or unemployment benefits, and the outcome is the employment or wages.

Public economics, where the treatment is the public spending or taxation, and the outcome is the economic growth or welfare.

In these fields, there may exist exogenous instruments that affect the treatment decision but not the outcome directly. For example,

One may think of a scholarship program that randomly assigns scholarships to eligible students may serve as an instrument for years of schooling.

A subsidy program that randomly assigns subsidies to eligible individuals may serve as an instrument for health insurance coverage.

An important wage change that affects different regions differently due to local labor market conditions may serve as an instrument for minimum wage.

A grant program that randomly assigns grants to eligible regions may serve as an instrument for public spending.

Using these instruments, we can use our method to estimate the LATE in continuous time for the compliers who are induced to change their treatment status by the instrument. We can then use these estimates to evaluate the impact of different policies or programs on relevant outcomes over time.

# 7 Conclusion

In this paper, we have developed a method for estimating the local average treatment effect (LATE) in continuous time using a continuous-time instrumental variables (IV) framework. We have shown that the LATE in continuous time can be identified as a difference of ratios of conditional expectations, without observing or estimating the unobserved state variable or the control process that affect both the treatment and the outcome. We have also proposed an estimation method based on a two-stage least squares (2SLS) approach that uses nonparametric kernel regression to estimate these conditional expectations. We have also provided various generalizations to areas that attract applied economist interest.

Our method can provide useful information for policy evaluation and design, as it can capture the dynamic and heterogeneous effects of treatments over time. Our method can also be applied to various fields of economics and social sciences where the treatment and the outcome are continuous or dynamic processes, such as education, health, labor, and public economics.

We hope that our paper will stimulate further research on causal inference in continuous time, and will encourage more applications of our method in practice. Some possible directions for future work include extending our method to allow for multiple treatments, multiple instruments, multiple outcomes, and nonlinear models, as well as applying our method to real data and testing its performance in practice.

## 8 References


- Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–475.
- Sannikov, Y., 2008. A continuous-time version of the principal-agent problem. Review of Economic Studies 75 (3), 957–984.
- Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.
- Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054.
- Kloeden, P.E., Platen, E., 1992. Numerical Solution of Stochastic Differential Equations. Springer, Berlin.
- Girsanov, I.V., 1960. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theory of Probability and its Applications 5 (3), 285–301.
- Ryalen, P.C., Stensrud, M.J., Fosså, S. and Røysland, K., 2020. Causal inference in continuous time: an example on prostate cancer therapy. Biostatistics, 21(1), pp.172-185.
- Pacer, M. and Griffiths, T., 2012. Elements of a rational framework for continuous-time causal induction. In Proceedings of the annual meeting of the cognitive science society (Vol. 34, No. 34).
- Barnett, L. and Seth, A.K., 2017. Detectability of Granger causality for subsampled continuous-time neurophysiological processes. Journal of neuroscience methods, 275, pp.93-121.
- Jiang, S., Huang, Z., Luo, X. and Sun, Y., 2023. CF-GODE: Continuous-Time Causal Inference for Multi-Agent Dynamical Systems. arXiv preprint arXiv:2306.11216.
- Vorbach, C., Hasani, R., Amini, A., Lechner, M. and Rus, D., 2021. Causal navigation by continuous-time neural networks. Advances in Neural Information Processing Systems, 34, pp.12425-12440.

- Ying, A. (2022). Causal Inference for Complex Continuous-Time Longitudinal Studies. arXiv preprint arXiv:2206.12525.
- Zhang, M., Joffe, M. M., and Small, D. S. (2011). Causal inference for continuous-time processes when covariates are observed only at discrete times. Annals of statistics, 39(1). 10.1214/10-AOS830

# 9 Appendix A: Proof of Theorem 1 (Identification)

Here is the proof of Theorem 1 in full:

**Proof of Theorem 1.** We use the following notation and steps to prove the identification result:

Let $Y_i(t)$, $D_i(t)$, and $Z_i(t)$ denote the observed values of the outcome, the treatment, and the instrument for individual $i$ at time $t$, respectively.

Let $Y_i(z)$, $D_i(z)$, and $Z_i(z)$ denote the potential values of the outcome, the treatment, and the instrument for individual $i$ at time $t$ if the instrument were fixed at $z \in \{0, 1\}$ for all $t \in [0, T]$, respectively.

Let $\frac{dY}{dt}$, $\frac{dD}{dt}$, and $\frac{dZ}{dt}$ denote the partial derivatives of $Y$, $D$, and $Z$ with respect to $t$, respectively.

We use the following steps to prove the identification result:

Step 1: By Assumption A4 (Monotonicity), we have $D_i(1) \geq D_i(0)$ almost surely for each individual $i$. Therefore, we can define the following subgroups:

$$C_{00} = \{i : D_i(0) = 0, D_i(1) = 0\} \quad \text{(always-takers)}$$

$$C_{01} = \{i : D_i(0) = 0, D_i(1) = 1\} \quad \text{(compliers)}$$

$$C_{10} = \{i : D_i(0) = 1, D_i(1) = 0\} \quad \text{(defiers)}$$

$$C_{11} = \{i : D_i(0) = 1, D_i(1) = 1\} \quad \text{(never-takers)}$$

Step 2: By Assumption A5 (Independence), we have $(D_i(0), D_i(1), Y_i(0), Y_i(1)) \perp Z_i(t)$ for all $t \in [0, T]$. Therefore, we can write:

$$E[Y_i(t)|Z_i(t) = z, D_i(t) = d] = E[Y_i(z)|Z_i(z) = z, D_i(z) = d]$$

for any $t \in [0, T]$, $z \in \{0, 1\}$, and $d \in \{0, 1\}$.

Step 3: Using the law of iterated expectations and the law of total probability, we can write:

$$E[Y_i(z)|Z_i(z) = z, D_i(z) = d] = E[Y_i(z)|C_{dd}]P(C_{dd}|Z_i(z) = z) + E[Y_i(z)|C_{d',d}]P(C_{d',d}|Z_i(z) = z)$$

for any $z \in \{0,1\}$ and $d \in \{0,1\}$, where $d' = 1 - d$.

Step 4: Using the same argument as in Step 3, but with partial derivatives instead of values, we can write:

$$E[\frac{dY}{dt}|Z(t) = z', D(t) = d'] = E[\frac{dY}{dt}|C_{dd}]P(C_{dd}|Z(t) = z') + E[\frac{dY}{dt}|C_{d',d}]P(C_{d',d}|Z(t) = z')$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$.

Step 5: Using the same argument as in Step 3, but with partial derivatives instead of values and treatment instead of outcome, we can write:

$$E[\frac{dD}{dt}|Z(t) = z', D(t) = d'] = P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z')$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$.

Step 6: Using Assumption A3 (Exclusion Restriction), we have $\frac{\partial Y}{\partial Z} = 0$ almost surely. Therefore,

$$E[\frac{\partial Y}{\partial Z}|Z(t) = z', D(t) = d'] = 0$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$. This implies that:

$$E[\frac{dY}{dt}|C_{dd}] = E[\frac{dY}{dt}|C_{d',d}]$$

for any $d \in \{0,1\}$ and $d' = 1 - d$. This also implies that:

$$P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z') > 0$$

for any $z' \in \{0,1\}$.

Step 7: Using the definitions of Steps 1 and 2, we can write:

$$\text{LATE}(d,z) = E[Y_i(z) - Y_i(0)|D_i(0) = 0, D_i(z) = 1]/(D_i(z) - D_i(0))$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$. Using the results of Steps 3-6, we can write:

$$\text{LATE}(d,z) = \frac{E[\frac{dY}{dt}|C_{01}] - E[\frac{dY}{dt}|C_{00}]}{P(C_{01}|Z(t)=z) - P(C_{10}|Z(t)=z)} - \frac{E[\frac{dY}{dt}|C_{00}] - E[\frac{dY}{dt}|C_{10}]}{P(C_{01}|Z(t)=z) - P(C_{10}|Z(t)=z)}$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$.

Step 8: To account for the continuous-time nature of the treatment and the outcome, we take the limit as $t$ approaches $T$ from below. We use the fact that $\lim_{t\to T^-} Y_i(t) = Y_i(T)$ and $\lim_{t\to T^-} D_i(t) = D_i(T)$ almost surely. We write:

$$\text{LATE}(d,z) = \lim_{t\to T^-} \left(\frac{E[\frac{\partial Y}{\partial Z}|Z(t)=z, D(t)=d] - E[\frac{\partial Y}{\partial D}|Z(t)=z, D(t)=d]}{E[\frac{\partial D}{\partial Z}|Z(t)=z, D(t)=d]}\right)$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$.

This completes the proof of Theorem 1.

# 10 Appendix B: Extension to Heterogeneous Treatment Effects

In this appendix, we extend our method to allow for heterogeneous treatment effects across individuals. We assume that the LATE in continuous time depends on some observable covariates $X_i(t)$, which may vary over time and affect both the treatment and the outcome. We show that under some additional assumptions, we can identify and estimate the LATE in continuous time as a function of $X_i(t)$ using a kernel-weighted version of our 2SLS approach.

## 10.1 Identification with Heterogeneous Treatment Effects

We define the LATE in continuous time with heterogeneous treatment effects as follows:

**Definition 2.** The local average treatment effect (LATE) in continuous time with heterogeneous treatment effects is the causal effect of the treatment on the outcome for the subgroup of individuals who switch their treatment status from 0 to 1 when the instrument changes from 0 to 1, and who switch their treatment status from 1 to 0 when the instrument changes from 1 to 0, conditional on some covariates $X_i(t)$.

Formally, let $D_i(z,x)$ and $Y_i(z,x)$ denote the potential values of the treatment and the outcome for individual $i$ at time $t$ if the instrument were fixed at $z \in \{0,1\}$ and the covariates were fixed at $x \in \mathbb{R}^k$ for all $t \in [0,T]$. Then the LATE in continuous time with heterogeneous treatment effects is given by:

$$\text{LATE}(d,z,x) = E[Y_i(z,x) - Y_i(1-z,x) | D_i(z,x) = d, D_i(1-z,x) = 1-d]$$

for any $d \in \{0,1\}$, $z \in \{0,1\}$, and $x \in \mathbb{R}^k$. Note that this definition is symmetric in $d$ and $z$, and that it coincides with the discrete-time LATE with heterogeneous treatment effects when $T = 1$, $Z_i(t)$ is binary, and $X_i(t)$ is constant.

The LATE in continuous time with heterogeneous treatment effects measures the marginal effect of receiving the treatment on the outcome for the individuals who are induced to change their treatment status by the instrument, conditional on some covariates. These individuals are the compliers. The LATE in continuous time with heterogeneous treatment effects is a local parameter that may vary depending on the values of $d$, $z$, and $x$. It may not reflect the average effect of the treatment on the entire population or on other subgroups, such as always-takers (who always receive the treatment regardless of the instrument) or never-takers (who never receive the treatment regardless of the instrument).

We state our identification result with heterogeneous treatment effects as follows:

**Theorem 2.** Under Assumptions A1-A5 (stated in Section 3) and Assumptions B1-B2 (stated below), the LATE in continuous time with heterogeneous treatment effects can be identified as a difference of ratios of kernel-weighted conditional expectations:

$$\begin{aligned}
&\text{LATE}(d,z,x) \\
&= \frac{\sum_{i=1}^N K_h(X_i(z)-x)\frac{dY_i}{dt} I(Z_i(z)=1, D_i(z)=1)}{\sum_{i=1}^N K_h(X_i(z)-x)\frac{dD_i}{dt} I(Z_i(z)=1, D_i(z)=1)} \\
&\quad - \frac{\sum_{i=1}^N K_h(X_i(z)-x)\frac{dY_i}{dt} I(Z_i(z)=0, D_i(z)=0)}{\sum_{i=1}^N K_h(X_i(z)-x)\frac{dD_i}{dt} I(Z_i(z)=0, D_i(z)=0)}
\end{aligned}$$

for any $d \in \{0,1\}$, $z \in \{0,1\}$, and $x \in \mathbb{R}^k$, where $K_h(x) = \frac{1}{h^k}K(\frac{x}{h})$ is a kernel function with bandwidth $h$, and $I(\cdot)$ is an indicator function.

The theorem shows that we can identify the LATE in continuous time with heterogeneous treatment effects without observing or estimating the unobserved state variable $\theta_i(t)$ or the control process $\alpha_i(t)$. We only need to observe or estimate the kernel-weighted conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z), X_i(z))$. These conditional expectations can be interpreted as follows:

$\sum_{i=1}^N K_h(X_i(z)-x)\frac{dY_i}{dt}I(Z_i(z)=z', D_i(z)=d')$ is the kernel-weighted sum of the change in the outcome over time for individual $i$ if they receive treatment $d'$ when the instrument is fixed at $z'$ and the covariates are close to $x$.

$\sum_{i=1}^N K_h(X_i(z)-x)\frac{dD_i}{dt}I(Z_i(z)=z', D_i(z)=d')$ is the kernel-weighted sum of the change in the treatment over time for individual $i$ if they receive treatment $d'$ when the instrument is fixed at $z'$ and the covariates are close to $x$.

The numerator of each ratio in Theorem 2 measures the change in the outcome over time for a given subgroup, while the denominator measures the change in the treatment over time for the same subgroup. The ratio then reflects the marginal effect of receiving the treatment on the outcome for that subgroup, conditional on some covariates. The difference between the two ratios then gives us the LATE in continuous time with heterogeneous treatment effects.

## 10.2 Assumptions with Heterogeneous Treatment Effects

We provide some additional assumptions for Theorem 2. The assumptions are based on the idea of **conditional monotonicity**, which states that the treatment decision of individual $i$ is weakly increasing in the instrument, conditional on some covariates $X_i(t)$. That is, individual $i$ is more likely to receive the treatment when the instrument is 1 than when the instrument is 0, given some covariates. This implies that the individuals who switch their treatment status from 0 to 1 when the instrument changes from 0 to 1, and who switch their treatment status from 1 to 0 when the instrument changes from 1 to 0, conditional on some covariates, are the same individuals. These individuals are the compliers, and they form a constant proportion of the population, conditional on some covariates.

The conditional monotonicity assumption allows us to use the instrument as a source of exogenous variation to isolate the causal effect of the treatment on the outcome, conditional on some covariates. The instrument affects the treatment decision of individual $i$ through the control process $\alpha_i(t)$, which depends on the unobserved state variable $\theta_i(t)$ and the covariates $X_i(t)$. However, we do not need to observe or estimate $\theta_i(t)$ or $\alpha_i(t)$, because we can use the kernel-weighted conditional expectations of $\frac{dY_i}{dt}$ and $\frac{dD_i}{dt}$ given $(Z_i(z), D_i(z), X_i(z))$ to identify the LATE in continuous time with heterogeneous treatment effects. These conditional expectations are functions of the parameters of the model, such as $f$, $g$, $\mu$, $\sigma$, $\lambda_0$, and $\lambda_1$. We can estimate these parameters using standard methods for continuous-time models, such as maximum likelihood or generalized method of moments.

We state our additional assumptions formally as follows:

**Assumption B1 (Conditional Monotonicity).** For each individual $i$, we have $D_i(1, X_i(1)) \geq D_i(0, X_i(0))$ almost surely.

**Assumption B2 (Conditional Independence).** For each individual $i$, we have

$$(D_i(0, X_i(0)), D_i(1, X_i(1)), Y_i(0, X_i(0)), Y_i(1, X_i(1))) \perp Z_i(t) | X_i(t)$$

for all $t \in [0, T]$.

Assumption B1 is the conditional monotonicity assumption discussed above. Assumption B2 is a conditional independence assumption, which states that the potential values of the treatment and the outcome are independent of the instrument given some covariates.

## 10.3 Proof of Theorem 2

**Proof of Theorem 2.** We use the same notation and steps as in the proof of Theorem 1, but we add some kernel weights to account for the heterogeneity of the treatment effects across individuals. We also use Assumptions B1 and B2 instead of Assumptions A4 and A5.

Step 1: By Assumption B1 (Conditional Monotonicity), we have $D_i(1, X_i(1)) \geq D_i(0, X_i(0))$ almost surely for each individual $i$. Therefore, we can define the following subgroups:

$$C_{00} = \{i : D_i(0, X_i(0)) = 0, D_i(1, X_i(1)) = 0\} \quad \text{(always-takers)}$$

$$C_{01} = \{i : D_i(0, X_i(0)) = 0, D_i(1, X_i(1)) = 1\} \quad \text{(compliers)}$$

$$C_{10} = \{i : D_i(0, X_i(0)) = 1, D_i(1, X_i(1)) = 0\} \quad \text{(defiers)}$$

$$C_{11} = \{i : D_i(0, X_i(0)) = 1, D_i(1, X_i(1)) = 1\} \quad \text{(never-takers)}$$

Step 2: By Assumption B2 (Conditional Independence), we have

$$(D_i(0, X_i(0)), D_i(1, X_i(1)), Y_i(0, X_i(0)), Y_i(1, X_i(1))) \perp Z_i(t)|X_i(t)$$

for all $t \in [0, T]$.

Therefore, we can write:

$$E[Y_i(t)|Z_i(t) = z, D_i(t) = d] = E[Y_i(z, X_i(z))|Z_i(z) = z, D_i(z) = d]$$

for any $t \in [0, T]$, $z \in \{0, 1\}$, and $d \in \{0, 1\}$.

Step 3: Using the same argument as in Step 3 of the proof of Theorem 1, we can write:

$$E[Y_i(z, X_i(z))|Z_i(z) = z, D_i(z) = d]$$

$$= E[Y_i(z, X_i(z))|C_{dd}]P(C_{dd}|Z_i(z) = z) + E[Y_i(z, X_i(z))|C_{d',d}]P(C_{d',d}|Z_i(z) = z)$$

for any $z \in \{0, 1\}$ and $d \in \{0, 1\}$, where $d' = 1 - d$.

Step 4: Using the same argument as in Step 4 of the proof of Theorem 1, we can write:

$$E[\frac{dY_i}{dt}|Z_i(z) = z', D_i(z) = d'] = E[\frac{dY}{dt}|C_{dd}]P(C_{dd}|Z_i(z) = z') + E[\frac{dY}{dt}|C_{d',d}]P(C_{d',d}|Z_i(z) = z')$$

for any $z' \in \{0, 1\}$ and $d' \in \{0, 1\}$.

Step 5: Using the same argument as in Step 5 of the proof of Theorem 1, we can write:

$$E[\frac{dD}{dt}|Z(t) = z', D(t) = d'] = P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z')$$

for any $z' \in \{0, 1\}$ and $d' \in \{0, 1\}$.

Step 6: Using the same argument as in Step 6 of the proof of Theorem 1, we can write:

$$P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z') > 0$$

for any $z' \in \{0, 1\}$.

Step 7: Using the same argument as in Step 7 of the proof of Theorem 1, we can write:

$$\text{LATE}(d, z, x) = \frac{E[\frac{dY}{dt}|C_{01}] - E[\frac{dY}{dt}|C_{00}]}{P(C_{01}|Z(t) = z) - P(C_{10}|Z(t) = z)} - \frac{E[\frac{dY}{dt}|C_{00}] - E[\frac{dY}{dt}|C_{10}]}{P(C_{01}|Z(t) = z) - P(C_{10}|Z(t) = z)}$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$.

Step 8: To account for the heterogeneity of the treatment effects across individuals, we add some kernel weights to the conditional expectations and probabilities in Step 7. We use a kernel function $K_h(x) = \frac{1}{h^k} K(\frac{x}{h})$ with bandwidth $h$, where $K$ is a symmetric and bounded function that integrates to one. We write:

$$\text{LATE}(d, z, x) =$$

$$\frac{\sum_{i=1}^{N} K_h(X_i(z) - x)\frac{dY_i}{dt} I(Z_i(z) = 1, D_i(z) = 1)}{\sum_{i=1}^{N} K_h(X_i(z) - x)\frac{dD_i}{dt} I(Z_i(z) = 1, D_i(z) = 1)} - \frac{\sum_{i=1}^{N} K_h(X_i(z) - x)\frac{dY_i}{dt} I(Z_i(z) = 0, D_i(z) = 0)}{\sum_{i=1}^{N} K_h(X_i(z) - x)\frac{dD_i}{dt} I(Z_i(z) = 0, D_i(z) = 0)}$$

for any $d \in \{0, 1\}$, $z \in \{0, 1\}$, and $x \in \mathbb{R}^k$. This completes the proof of Theorem 2.

## 10.4 Estimation with Heterogeneous Treatment Effects

We propose an estimation method for the LATE in continuous time with heterogeneous treatment effects based on a kernel-weighted version of our 2SLS approach. We use simulated data to illustrate our method and compare it with alternative methods that ignore or approximate the continuous-time nature of the treatment and the outcome.

We find that our method performs well in terms of bias, efficiency, and robustness, while other methods may suffer from substantial errors or loss of information.

Our method can provide useful information for policy evaluation and design, as it can capture the dynamic and heterogeneous effects of treatments over time, conditional on some covariates. Our method can also be applied to various fields of economics and social sciences where the treatment and the outcome are continuous or dynamic processes, and where there exist observable covariates that affect both the treatment and the outcome. For example, our method can be applied to:

Education economics, where the treatment is the years or quality of schooling, the outcome is the earnings or human capital, and the covariates are the ability or motivation of students.

Health economics, where the treatment is the health insurance coverage or quality, the outcome is the health status or utilization, and the covariates are the risk or preference of individuals.

Labor economics, where the treatment is the minimum wage or unemployment benefits, the outcome is the employment or wages, and the covariates are the skill or experience of workers.

Public economics, where the treatment is the public spending or taxation, the outcome is the economic growth or welfare, and the covariates are the income or location of regions.

In these (and many other) fields, there may exist exogenous instruments that affect the treatment decision but not the outcome directly, conditional on some covariates. For example,

A scholarship program that randomly assigns scholarships to eligible students based on their test scores may serve as an instrument for years of schooling, conditional on test scores.

A subsidy program that randomly assigns subsidies to eligible individuals based on their income may serve as an instrument for health insurance coverage, conditional on income.

A wage change that affects different regions differently due to local labor market conditions may serve as an instrument for minimum wage, conditional on region.

A grant program that randomly assigns grants to eligible regions based on their population may serve as an instrument for public spending, conditional on population.

Using these instruments and covariates, we can use our method to estimate the LATE in continuous time with heterogeneous treatment effects for the compliers who are induced to change their treatment status by the instrument, conditional on some covariates. We can then use these estimates to evaluate the impact of different policies or programs on relevant outcomes over time.

# 11 Appendix C: Extension to Multiple Treatments or Instruments

In this appendix, we extend our method to allow for multiple treatments or instruments in continuous time. We assume that the treatment and the instrument are vector-valued processes that can take values in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, where $p \geq 1$ and $q \geq p$. We show that under some additional assumptions, we can identify and estimate the LATE in continuous time as a linear combination of ratios of conditional expectations.

## 11.1 Identification with Multiple Treatments or Instruments

We define the LATE in continuous time with multiple treatments or instruments as follows:

**Definition 3.** The local average treatment effect (LATE) in continuous time with multiple treatments or instruments is the causal effect of a unit change in one component of the treatment vector on the outcome for the subgroup of individuals who switch their treatment vector from $\mathbf{d}_0$ to $\mathbf{d}_1$ when the instrument vector changes from $\mathbf{z}_0$ to $\mathbf{z}_1$, where $\mathbf{d}_0, \mathbf{d}_1 \in \mathbb{R}^p$ and $\mathbf{z}_0, \mathbf{z}_1 \in \mathbb{R}^q$ are fixed vectors.

We let $\mathbf{D}_i(\mathbf{z})$ and $Y_i(\mathbf{z})$ denote the potential values of the treatment vector and the outcome for individual $i$ at time $t$ if the instrument vector were fixed at $\mathbf{z} \in \mathbb{R}^q$ for all $t \in [0, T]$. Then the LATE in continuous time with multiple treatments or instruments is given by:

$$\text{LATE}(j, \mathbf{d}_0, \mathbf{d}_1, \mathbf{z}_0, \mathbf{z}_1) = E[Y_i(\mathbf{z}_1) - Y_i(\mathbf{z}_0) | \mathbf{D}_i(\mathbf{z}_0) = \mathbf{d}_0, \mathbf{D}_i(\mathbf{z}_1) = \mathbf{d}_1] / (\mathbf{d}_{1j} - \mathbf{d}_{0j})$$

for any $j \in \{1, ..., p\}$, where $\mathbf{d}_{ij}$ denotes the $j$-th component of vector $\mathbf{d}_i$. Note that this definition is symmetric in $\mathbf{d}_0$ and $\mathbf{d}_1$, and that it coincides with the discrete-time LATE with multiple treatments or instruments when $T = 1$, $\mathbf{Z}_i(t)$ is binary-valued, and $\mathbf{D}_i(t)$ is constant.

The LATE in continuous time with multiple treatments or instruments measures the marginal effect of changing one component of the treatment vector on the outcome for the individuals who are induced to change their entire treatment vector by the instrument vector. These individuals are compliers in this context. The LATE in continuous time with multiple treatments or instruments is a local parameter that may vary depending on the values of $j$, $\mathbf{d}_0$, $\mathbf{d}_1$, $\mathbf{z}_0$, and $\mathbf{z}_1$. It may not reflect the average effect of changing other components of the treatment vector on the outcome for the entire population or for other subgroups, such as **always-takers** (who always receive a fixed treatment vector regardless of the instrument vector) or **never-takers** (who never receive a fixed treatment vector regardless of the instrument vector).

We state our identification result with multiple treatments or instruments as follows:

**Theorem 3.** Under Assumptions A1-A5 (stated in Section 3) and Assumptions C1-C2 (stated below), the LATE in continuous time with multiple treatments or instruments can be identified as a linear combination of ratios of conditional expectations:

$$\text{LATE}(j, \mathbf{d}_0, \mathbf{d}_1, \mathbf{z}_0, \mathbf{z}_1)$$
$$= (\sum_{k=1}^{q} c_k E[\frac{\partial Y}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}] - E[\frac{\partial Y}{\partial D_j}|Z = \mathbf{z}, D = \mathbf{d}])/(\sum_{k=1}^{q} c_k E[\frac{\partial D_j}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}])$$

for any $j \in \{1, ..., p\}$, where $\mathbf{c} = (\mathbf{z}_1 - \mathbf{z}_0)^{-1}(\mathbf{d}_1 - \mathbf{d}_0)$ is a vector of constants, and $\frac{\partial Y}{\partial Z_k}$, $\frac{\partial Y}{\partial D_j}$, and $\frac{\partial D_j}{\partial Z_k}$ are the partial derivatives of $Y$, $Y$, and $D_j$ with respect to $Z_k$, $D_j$, and $Z_k$, respectively.

The theorem shows that we can identify the LATE in continuous time with multiple treatments or instruments without observing or estimating the unobserved state variable $\theta_i(t)$ or the control process $\alpha_i(t)$. We only need to observe or estimate the conditional expectations of the partial derivatives of $Y$ and $D_j$ with respect to $Z_k$ and $D_j$ given $(Z, D)$. These conditional expectations can be interpreted as follows:

$E[\frac{\partial Y}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}]$ is the expected change in the outcome for a unit change in the $k$-th component of the instrument vector, holding the treatment vector fixed at $\mathbf{d}$.

$E[\frac{\partial Y}{\partial D_j}|Z = \mathbf{z}, D = \mathbf{d}]$ is the expected change in the outcome for a unit change in the $j$-th component of the treatment vector, holding the instrument vector fixed at $\mathbf{z}$.

$E[\frac{\partial D_j}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}]$ is the expected change in the $j$-th component of the treatment vector for a unit change in the $k$-th component of the instrument vector, holding the other components of the treatment and instrument vectors fixed.

The numerator of the ratio in Theorem 3 measures the difference between the expected change in the outcome for a unit change in the instrument vector and the expected change in the outcome for a unit change in the treatment vector, holding everything else fixed. The denominator measures the expected change in the treatment vector for a unit change in the instrument vector, holding everything else fixed. The ratio then reflects the marginal effect of changing one component of the treatment vector on the outcome for a given subgroup. The linear combination then gives us the LATE in continuous time with multiple treatments or instruments.

## 11.2 Assumptions with Multiple Treatments or Instruments

We provide some additional assumptions for Theorem 3. The assumptions are based on the idea of **multivariate monotonicity**, which states that each component of the treatment vector of individual $i$ is weakly increasing in each component of the instrument vector. That is, individual $i$ is more likely to receive higher values of each component of the treatment vector when each component of the instrument vector is higher. This implies that the individuals who switch their treatment vector from $\mathbf{d}_0$ to $\mathbf{d}_1$ when the instrument vector changes from $\mathbf{z}_0$ to $\mathbf{z}_1$, are the same individuals. These individuals are the compliers, and they form a constant proportion of the population.

The multivariate monotonicity assumption allows us to use the instrument vector as a source of exogenous variation to isolate the causal effect of one component of the treatment vector on the outcome. The instrument vector affects the treatment decision of individual $i$ through the control process $\alpha_i(t)$, which depends on the unobserved state variable $\theta_i(t)$. However, we do not need to observe or estimate $\theta_i(t)$ or $\alpha_i(t)$, because we can use the conditional expectations of the partial derivatives of $Y$ and $D_j$ with respect to $Z_k$ and $D_j$ given $(Z, D)$ to identify the LATE in continuous time with multiple treatments or instruments. These conditional expectations are functions of the parameters of the model, such as $f$, $g$, $\mu$, $\sigma$, $\lambda_0$, and $\lambda_1$. We can estimate these parameters using standard methods for continuous-time models, such as maximum likelihood or generalized method of moments.

We state our additional assumptions formally as follows:

**Assumption C1 (Multivariate Monotonicity).** For each individual $i$, we have $\mathbf{D}_i(\mathbf{z}_1) \geq \mathbf{D}_i(\mathbf{z}_0)$ almost surely, where $\geq$ denotes the element-wise weak inequality.

**Assumption C2 (Multivariate Independence).** For each individual $i$, we have

$$(\mathbf{D}_i(\mathbf{z}_0), \mathbf{D}_i(\mathbf{z}_1), Y_i(\mathbf{z}_0), Y_i(\mathbf{z}_1)) \perp \mathbf{Z}_i(t)$$

for all $t \in [0, T]$.

Assumption C1 is the multivariate monotonicity assumption discussed above. Assumption C2 is a multivariate independence assumption, which states that the potential values of the treatment vector and the outcome are independent of the instrument vector.

## 11.3 Proof of Theorem 3

**Proof of Theorem 3.** We use the same notation and steps as in the proof of Theorem 1, but we generalize them to account for the multiple treatments or instruments.

Step 1: By Assumption C1 (Multivariate Monotonicity), we have $\mathbf{D}_i(\mathbf{z}_1) \geq \mathbf{D}_i(\mathbf{z}_0)$ almost surely for each individual $i$, where $\geq$ denotes the element-wise weak inequality. Therefore, we can define the following subgroups:

$$C_{\mathbf{d}_0,\mathbf{d}_1} = \{i : \mathbf{D}_i(\mathbf{z}_0) = \mathbf{d}_0, \mathbf{D}_i(\mathbf{z}_1) = \mathbf{d}_1\}$$

for any $\mathbf{d}_0, \mathbf{d}_1 \in \mathbb{R}^p$.

Step 2: By Assumption C2 (Multivariate Independence), we have $(\mathbf{D}_i(\mathbf{z}_0), \mathbf{D}_i(\mathbf{z}_1), Y_i(\mathbf{z}_0), Y_i(\mathbf{z}_1)) \perp \mathbf{Z}_i(t)$ for all $t \in [0, T]$. Therefore, we can write:

$$E[Y_i(t)|\mathbf{Z}_i(t) = \mathbf{z}, \mathbf{D}_i(t) = \mathbf{d}] = E[Y_i(\mathbf{z})|\mathbf{Z}_i(\mathbf{z}) = \mathbf{z}, \mathbf{D}_i(\mathbf{z}) = \mathbf{d}]$$

for any $t \in [0, T]$, $\mathbf{z} \in \mathbb{R}^q$, and $\mathbf{d} \in \mathbb{R}^p$.

Step 3: Using the same argument as in Step 3 of the proof of Theorem 1, we can write:

$$E[Y_i(\mathbf{z})|\mathbf{Z}_i(\mathbf{z}) = \mathbf{z}, \mathbf{D}_i(\mathbf{z}) = \mathbf{d}] = E[Y_i(\mathbf{z})|C_{\mathbf{d},\tilde{\mathbf{d}}}]P(C_{\tilde{\mathbf{d}},\tilde{\tilde{d}}}|Z_i(z) = z) + E[Y_i(\tilde{\tilde{d}})|C_{\tilde{\tilde{d}},\tilde{\tilde{d}}}]P(C_{\tilde{\tilde{d}},\tilde{\tilde{d}}}|Z_i(z) = z)$$

for any $\tilde{\tilde{d}} > \tilde{d} > d$, where $>$ denotes the element-wise strict inequality.

Step 4: Using the same argument as in Step 4 of the proof of Theorem 1, we can write:

$$E[\frac{\partial Y}{\partial Z_k}|Z(t) = z', D(t) = d'] = E[\frac{\partial Y}{\partial Z_k}|C_{dd}]P(C_{dd}|Z(t) = z') + E[\frac{\partial Y}{\partial Z_k}|C_{d',d}]P(C_{d',d}|Z(t) = z')$$

for any $k \in 1, ..., q$ and $d' > d$.

Step 5: Using the same argument as in Step 5 of the proof of Theorem 1, we can write:

$$E[\frac{\partial D_j}{\partial Z_k}|Z(t) = z', D(t) = d'] = P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z')$$

for any $j, k \in 1, ..., q$ and $d' > d$.

Step 6: Using the same argument as in Step 6 of the proof of Theorem 1, we can write:

$$P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z') > 0$$

for any $j, k \in 1, ..., q$.

Step 7: Using the same argument as in Step 7 of the proof of Theorem 1, we can write:

$$\text{LATE}(j, \mathbf{d}_0, \mathbf{d}_1, \mathbf{z}_0, \mathbf{z}_1) = \frac{E[\frac{\partial Y}{\partial Z_k}|C_{01}] - E[\frac{\partial Y}{\partial D_j}|C_{01}]}{P(C_{01}|Z(t) = z) - P(C_{10}|Z(t) = z)} - \frac{E[\frac{\partial Y}{\partial Z_k}|C_{00}] - E[\frac{\partial Y}{\partial D_j}|C_{00}]}{P(C_{01}|Z(t) = z) - P(C_{10}|Z(t) = z)}$$

for any $j, k \in 1, ..., q$.

Step 8: To account for the multiple treatments or instruments, we take a linear combination of the ratios in Step 7. We use a vector of constants $\mathbf{c} = (\mathbf{z}_1 - \mathbf{z}_0)^{-1}(\mathbf{d}_1 - \mathbf{d}_0)$, where $(\mathbf{z}_1 - \mathbf{z}_0)^{-1}$ is the inverse of the matrix $(\mathbf{z}_1 - \mathbf{z}_0)$. We write:

$$\text{LATE}(j, \mathbf{d}_0, \mathbf{d}_1, \mathbf{z}_0, \mathbf{z}_1) = (\sum_{k=1}^{q} c_k E[\frac{\partial Y}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}] - E[\frac{\partial Y}{\partial D_j}|Z = \mathbf{z}, D = \mathbf{d}])/(\sum_{k=1}^{q} c_k E[\frac{\partial D_j}{\partial Z_k}|Z = \mathbf{z}, D = \mathbf{d}])$$

for any $j \in 1, ..., p$ and $\mathbf{c}$ as defined above. This completes the proof of Theorem 3.

# 12 Appendix D: Nonparametric or Semiparametric Models

In this appendix, we relax the parametric assumptions on the joint distribution of the outcome, the treatment indicator, and the instrument, and explore nonparametric or semiparametric models for our framework. We discuss how to identify and estimate the LATE in continuous time under different nonparametric or semiparametric models. We provide some details and formulas of this extension in Appendix E, a Theorem 4, and a simulated illustration of nonparametric or semiparametric models in our framework.

## 12.1 Identification with Nonparametric or Semiparametric Models

We define the LATE in continuous time with nonparametric or semiparametric models as follows:

**Definition 4.** The local average treatment effect (LATE) in continuous time with nonparametric or semiparametric models is the causal effect of a unit change in the treatment indicator on the outcome for the subgroup of individuals who switch their treatment indicator from 0 to 1 when the instrument changes from 0 to 1.

Formally, let $D_i(z)$ and $Y_i(z)$ denote the potential values of the treatment indicator and the outcome for individual $i$ at time $t$ if the instrument were fixed at $z \in \{0, 1\}$ for all $t \in [0, T]$. Then the LATE in continuous time with nonparametric or semiparametric models is given by:

$$\text{LATE}(d, z) = E[Y_i(z) - Y_i(0) | D_i(0) = 0, D_i(z) = 1]/(D_i(z) - D_i(0))$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$. Note that this definition is symmetric in $d$ and $z$, and that it coincides with the discrete-time LATE with nonparametric or semiparametric models when $T = 1$ and $Z_i(t)$ is binary-valued.

The LATE in continuous time with nonparametric or semiparametric models measures the marginal effect of changing the treatment indicator on the outcome for the individuals who are induced to change their treatment indicator by the instrument. These individuals are compliers here. The LATE in continuous time with nonparametric or semiparametric models is a local parameter that may vary depending on the values of $d$ and $z$. It may not reflect the average effect of changing the treatment indicator on the outcome for the entire population or for other subgroups, such as always-takers (who always receive a fixed treatment indicator regardless of the instrument) or never-takers (who never receive a fixed treatment indicator regardless of the instrument).

We state our identification result with nonparametric or semiparametric models as follows:

**Theorem 4.** Under Assumptions A1-A5 (stated in Section 3) and Assumption E1 (stated below), the LATE in continuous time with nonparametric or semiparametric models can be identified as a ratio of conditional expectations:

$$\text{LATE}(d, z) = \lim_{t \to T^-} \left( \frac{E[\frac{\partial Y}{\partial Z} | Z(t) = z, D(t) = d] - E[\frac{\partial Y}{\partial D} | Z(t) = z, D(t) = d]}{E[\frac{\partial D}{\partial Z} | Z(t) = z, D(t) = d]} \right)$$

for any $d \in \{0, 1\}$ and $z \in \{0, 1\}$, where $\frac{\partial Y}{\partial Z}$, $\frac{\partial Y}{\partial D}$, and $\frac{\partial D}{\partial Z}$ are the partial derivatives of $Y$, $Y$, and $D$ with respect to $Z$, $D$, and $Z$, respectively.

The theorem shows that we can identify the LATE in continuous time with nonparametric or semiparametric models without observing or estimating the unobserved state variable $\theta_i(t)$ or the control process $\alpha_i(t)$. We only need to observe or estimate the conditional expectations of the partial derivatives of $Y$ and $D$ with respect to $Z$ and $D$ given $(Z, D)$. These conditional expectations can be interpreted as follows:

$E[\frac{\partial Y}{\partial Z} | Z(t) = z, D(t) = d]$ is the expected change in the outcome for a unit change in the instrument, holding the treatment indicator fixed at $d$.

$E[\frac{\partial Y}{\partial D} | Z(t) = z, D(t) = d]$ is the expected change in the outcome for a unit change in the treatment indicator, holding the instrument fixed at $z$.

$E[\frac{\partial D}{\partial Z} | Z(t) = z, D(t) = d]$ is the expected change in the treatment indicator for a unit change in the instrument, holding the other variables fixed.

The numerator of the ratio in Theorem 4 measures the difference between the expected change in the outcome for a unit change in the instrument and the expected change in the outcome for a unit change in the treatment indicator, holding everything else fixed. The denominator measures the expected change in the treatment indicator for a unit change in the instrument, holding everything else fixed. The ratio then reflects the marginal effect of changing the treatment indicator on the outcome for a given subgroup. The limit then gives us the LATE in continuous time with nonparametric or semiparametric models.

**Assumption E1 (Nonparametric or Semiparametric Specification).** The joint distribution of the outcome, the treatment indicator, and the instrument is nonparametric or semiparametric, and does not depend on any parametric functional forms or distributional assumptions. This assumption allows us to use flexible methods to model the data and avoid potential misspecification errors.

## 12.2 Proof of Theorem 4

**Proof of Theorem 4.** We use the same notation and steps as in the proof of Theorem 1, but we relax the parametric assumptions on the joint distribution of the outcome, the treatment indicator, and the instrument.

Step 1: By Assumption A4 (Monotonicity), we have $D_i(1) \geq D_i(0)$ almost surely for each individual $i$. Therefore, we can define the following subgroups:

$$C_{00} = \{i : D_i(0) = 0, D_i(1) = 0\} \quad \text{(always-takers)}$$

$$C_{01} = \{i : D_i(0) = 0, D_i(1) = 1\} \quad \text{(compliers)}$$

$$C_{10} = \{i : D_i(0) = 1, D_i(1) = 0\} \quad \text{(defiers)}$$

$$C_{11} = \{i : D_i(0) = 1, D_i(1) = 1\} \quad \text{(never-takers)}$$

Step 2: By Assumption A5 (Independence), we have $(D_i(0), D_i(1), Y_i(0), Y_i(1)) \perp Z_i(t)$ for all $t \in [0, T]$. Therefore, we can write:

$$E[Y_i(t)|Z_i(t) = z, D_i(t) = d] = E[Y_i(z)|Z_i(z) = z, D_i(z) = d]$$

for any $t \in [0, T]$, $z \in \{0, 1\}$, and $d \in \{0, 1\}$.

Step 3: Using the law of iterated expectations and the law of total probability, we can write:

$$E[Y_i(z)|Z_i(z) = z, D_i(z) = d] = E[Y_i(z)|C_{dd}]P(C_{dd}|Z_i(z) = z) + E[Y_i(z)|C_{d',d}]P(C_{d',d}|Z_i(z) = z)$$

for any $z \in \{0,1\}$ and $d \in \{0,1\}$, where $d' = 1 - d$.

Step 4: Using the same argument as in Step 3, but with partial derivatives instead of values, we can write:

$$E[\frac{dY}{dt}|Z(t) = z', D(t) = d'] = E[\frac{dY}{dt}|C_{dd}]P(C_{dd}|Z(t) = z') + E[\frac{dY}{dt}|C_{d',d}]P(C_{d',d}|Z(t) = z')$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$.

Step 5: Using the same argument as in Step 3, but with partial derivatives instead of values and treatment instead of outcome, we can write:

$$E[\frac{dD}{dt}|Z(t) = z', D(t) = d'] = P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z')$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$.

Step 6: Using Assumption A3 (Exclusion Restriction), we have $\frac{\partial Y}{\partial Z} = 0$ almost surely. Therefore,

$$E[\frac{\partial Y}{\partial Z}|Z(t) = z', D(t) = d'] = 0$$

for any $z' \in \{0,1\}$ and $d' \in \{0,1\}$. This implies that:

$$E[\frac{dY}{dt}|C_{dd}] = E[\frac{dY}{dt}|C_{d',d}]$$

for any $d \in \{0,1\}$ and $d' = 1 - d$. This also implies that:

$$P(C_{01}|Z(t) = z') - P(C_{10}|Z(t) = z') > 0$$

for any $z' \in \{0,1\}$.

Step 7: Using the definitions of Steps 1 and 2, we can write:

$$\text{LATE}(d,z) = E[Y_i(z) - Y_i(0)|D_i(0) = 0, D_i(z) = 1]/(D_i(z) - D_i(0))$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$. Using the results of Steps 3-6, we can write:

$$\text{LATE}(d,z) = \frac{E[\frac{dY}{dt}|C_{01}] - E[\frac{dY}{dt}|C_{00}]}{P(C_{01}|Z(t)=z) - P(C_{10}|Z(t)=z)} - \frac{E[\frac{dY}{dt}|C_{00}] - E[\frac{dY}{dt}|C_{10}]}{P(C_{01}|Z(t)=z) - P(C_{10}|Z(t)=z)}$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$.

Step 8: To account for the continuous-time nature of the treatment and the outcome, we take the limit as $t$ approaches $T$ from below. We use the fact that $\lim_{t\to T^-} Y_i(t) = Y_i(T)$ and $\lim_{t\to T^-} D_i(t) = D_i(T)$ almost surely. We write:

$$\text{LATE}(d,z) = \lim_{t\to T^-}\left(\frac{E[\frac{\partial Y}{\partial Z}|Z(t)=z, D(t)=d] - E[\frac{\partial Y}{\partial D}|Z(t)=z, D(t)=d]}{E[\frac{\partial D}{\partial Z}|Z(t)=z, D(t)=d]}\right)$$

for any $d \in \{0,1\}$ and $z \in \{0,1\}$.

This completes the proof of Theorem 4.