Estimating Treatment Effects from Irregular Time Series Observations with Hidden Confounders

Defu Cao[*1]**, James Enouen**[*1]**, Yujing Wang**[2]**, Xiangchen Song**[3]**,**
Chuizheng Meng[1]**, Hao Niu**[4]**, Yan Liu**[1]

[1]University of Southern California
[2]Peking University
[3]Carnegie Mellon University
[4]KDDI Research, Inc.
{defucao, enouen, chuizhem, yanliu.cs}@usc.edu, yujwang@pku.edu.cn
xiangchensong@cmu.edu, ha-niu@kddi.com

Abstract

Causal analysis for time series data, in particular estimating individualized treatment effect (ITE), is a key task in many real-world applications, such as finance, retail, healthcare, etc. Real-world time series can include large-scale, irregular, and intermittent time series observations, raising significant challenges to existing work attempting to estimate treatment effects. Specifically, the existence of hidden confounders can lead to biased treatment estimates and complicate the causal inference process. In particular, anomaly hidden confounders which exceed the typical range can lead to high variance estimates. Moreover, in continuous time settings with irregular samples, it is challenging to directly handle the dynamics of causality. In this paper, we leverage recent advances in Lipschitz regularization and neural controlled differential equations (CDE) to develop an effective and scalable solution, namely LipCDE, to address the above challenges. LipCDE can directly model the dynamic causal relationships between historical data and outcomes with irregular samples by considering the boundary of hidden confounders given by Lipschitz constrained neural networks. Furthermore, we conduct extensive experiments on both synthetic and real-world datasets to demonstrate the effectiveness and scalability of LipCDE.

Introduction

Estimating individualized treatment effects (ITE) for time series data, which makes predictions about causal effects of actions (Zhang, Cao, and Liu 2022), is one key task in many domains, including marketing (Brodersen et al. 2015; Abadie, Diamond, and Hainmueller 2010), education (Mandel et al. 2014), healthcare (Kuzmanovic, Hatt, and Feuerriegel 2021), etc. However, the existence of confounders can introduce bias into the estimation (Simpson 1951; Pearl et al. 2000). For example, in finance applications, multi-factor investing strategies can give investors a deeper understanding of the risk drivers underlying a portfolio. The unobserved factors (i.e., hidden confounders), which typically happen at irregular time stamps and are not reflected in finance system records or are difficult to observe, could bring bias by influencing

both interventions and stock returns. The reason is that even a small number of existing factors (such as Small Minus Big and High Minus Low) could significantly explain the cross-section of stock returns (D'Acunto et al. 2021). If we can simulate such hidden confounders within a reasonable range, we are able to obtain treatment estimates with reduced bias and variance by making appropriate impact assumptions on the relationship between treatments and outcomes (Wang and Blei 2019).

Estimating ITE is an extremely challenging task in continuous time settings with hidden confounders. First, estimating treatment effects in large-scale irregular and sparse time series still has considerable room for improvement as previous works fail to consider the continuous time setting, where it is difficult to handle the dynamic behavior and complex interactions of covariates and treatments (Gao et al. 2021). Second, hidden confounders' values generated by randomness and noise can introduce high variance and undesirable explanations. For example, in healthcare applications, according to domain knowledge of drug resistance, the response to single-agent immune-checkpoint inhibitors (ICI) in uremic patients ranged from 15% to 31% (Zibelman, Ramamurthy, and Plimack 2016). Consequently, when left unconsidered, drug resistance will introduce biased estimates of treatment effects. Furthermore, any substitute confounders generated by data-driven methods with an impact on outcomes over 31% can lead to high variance.

Recently, there have been several attempts to address these challenges. To model hidden confounders over time, (Bahadori and Heckerman 2021) introduce a new causal prior graph for the confounding information and concept completeness to improve the interpretability of prediction models; (Mastakouri, Schölkopf, and Janzing 2021) study the identification of direct and indirect causes for causal feature selection in time series solely based on observational data. Deconfounding-based models (Hatt and Feuerriegel 2021; Bica, Alaa, and Van Der Schaar 2020) use latent variables given by their factor model as substitutes for the hidden confounders to render the assigned treatments conditionally independent. However, existing works either cannot handle irregular time series (Bahadori and Heckerman 2021; Mas-

---

takouri, Schölkopf, and Janzing 2021), or have strong assumptions (Hatt and Feuerriegel 2021; Bica, Alaa, and Van Der Schaar 2020). Furthermore, the range of hidden confounders generated by previous data-driven works is possibly unjustifiable, which will distort (obscure or augment) the true causal relationship between treatments and outcomes.

In this work, we consider the task of estimating treatment effects under continuous time settings with multi-cause hidden confounders (which affect multiple treatments and the outcome). To tackle the above two challenges, we propose a novel Lipschitz regularized neural controlled differential equation (LipCDE) model for estimation by obtaining the constrained time-varying hidden confounders. Specifically, LipCDE first infers the interrelationship of hidden confounders on treatment by estimating the boundary of hidden confounders: we decompose the historical covariates into low-frequency components and high-frequency components in the spectral domain. Then we use Lipschitz regularization (Araujo et al. 2021) on the decomposition to get the latent representation. Afterward, we model the historical trajectories with neural CDE using sparse numerical solvers, which is one of the most suitable methods for large-scale problems under the continuous time setting (Fröhlich, Loos, and Hasenauer 2019). In this way, we can explicitly model the observed irregular sequential data as a process evolving continuously in time with a dynamic causal relationship to equip the LipCDE with interpretability. In the outcome model, we re-weight the population of all participating patients and balance the representation via applying the inverse probability of treatment weighting (IPTW) strategy (Lim, Alaa, and van der Schaar 2018).

In this paper, we conduct extensive experiments on both simulated and real-world datasets. Experimental results show that LipCDE outperforms other state-of-the-art estimating treatment effect approaches. From a qualitative perspective, experiments show that LipCDE is in agreement with the true underlying hidden confounders in simulated environments, which can effectively eliminate bias in causal models (Pearl et al. 2000). In addition, the average RMSE of TSD (Bica, Alaa, and Van Der Schaar 2020) and SeqConf (Hatt and Feuerriegel 2021) on MIMIC-III's blood pressure outcome and COVID-19 datasets decreases by 28.7% and 32.3%, respectively. To the best of our knowledge, this is the first complete estimating treatment effects model that considers both the boundary of hidden confounders and the continuous time setting.

We summarize the main **contributions** as follows:

- LipCDE utilizes a convolutional operation with Lipschitz regularization on the spectral domain and neural controlled differential equation from observed data to obtain hidden confounders, which are bounded to reduce the high variance of treatment effect estimation.

- LipCDE can fully use information of observed data and dynamic time intervals, allowing the continuous inclusion of input interventions and supporting irregularly sampled time series.

- Sufficient experiments demonstrate the effectiveness of LipCDE in estimating treatment effect on both synthetic and real-world datasets. Particularly, experiments on MIMIC-III and COVID-19 demonstrate the potential of LipCDE for health care applications in personalized medical recommendation.

Related Work

Treatment effects learning in the static setting. In recent years, there has been a significant increase in interest in the study of causal inference accomplished through representational learning (Kallus, Mao, and Udell 2018; Curth and van der Schaar 2021). (Johansson, Shalit, and Sontag 2016) propose to take advantage of the multiple processing methods assigned in a static environment. (Shalit, Johansson, and Sontag 2017) show that balancing the representational distributions of the treatment and control groups can help upper limits of error for counterfactual outcome estimates. However, these approaches rely on the strong ignorability assumption, which ignores the influence of implicit hidden confounders. Many works focus on relaxing such assumptions with the consideration of hidden confounders including domain adversarial training (Berrevoets et al. 2020; Curth and van der Schaar 2021). (Guo, Li, and Liu 2020a) and (Guo, Li, and Liu 2020b) propose to unravel the patterns of hidden confounders from the network structure and observed features by learning the representations of hidden confounders and using the representations for potential outcome prediction. (Wang and Blei 2019) propose to estimate confounding factors in a static setting using a latent factor model and then infer potential outcomes using bias adjustment. Nevertheless, such works fail to take advantage of the dynamic evolution of the observed variables and the inter-individual relationships which are present in the time-dynamic setting.

Treatment effects learning in the dynamic setting without hidden confounders. There are many related previous works estimating treatment effects in dynamic settings including g-computation formula, g-estimation of structural nested mean models (Hernán and Robins 2010), IPTW in marginal structural models (MSMs) (Robins and Hernán 2009), and recurrent marginal structural networks (RMSNs) (Lim, Alaa, and van der Schaar 2018), CRN (Bica et al. 2020) etc. In addition, Gaussian processes (Schulam and Saria 2017) and bayesian nonparametrics (Roy, Lum, and Daniels 2017) have been tailored to estimate treatment response in a continuous time setting in order to incorporate non-deterministic quantification. Besides, (Soleimani, Subbaswamy, and Saria 2017) relies on regularization to decompose the observed data into shared and signal-specific components in treatment response curves from multivariate longitudinal data. However, those models still need constraint methods to guarantee the posterior consistency of the sub-component modules and cannot directly model the dynamic causal relationship between different time intervals. While (Seedat et al. 2022; De Brouwer, Gonzalez, and Hyland 2022) directly model the dynamic causal relationship, they make a strong assumption with no hidden confounders, which does not have the flexibility to be applied to all real-world scenarios.

Treatment effect learning in the dynamic setting with hidden confounders. Rather than making strong ignorability assumptions, (Pearl 2012) and (Kuroki and Pearl 2014)

theoretically prove that observed proxy variables can be used to capture hidden confounders and estimate treatment effects. (Veitch, Sridhar, and Blei 2020) use network information as a proxy variable to mitigate confounding bias without utilizing the characteristics of the instances. TSD (Bica, Alaa, and Van Der Schaar 2020) introduces recurrent neural networks in the factor model to estimate the dynamics of confounders. In a similar vein, (Hatt and Feuerriegel 2021) propose a sequential deconfounder to infer hidden confounders by using Gaussian process latent variable model and DTA (Kuzmanovic, Hatt, and Feuerriegel 2021) estimates treatment effects under dynamic setting using observed data as noisy proxies. Besides, DSW (Liu, Yin, and Zhang 2020) infers the hidden confounders by using a deep recursive weighted neural network that combines current treatment assignment and historical information. DNDC (Ma et al. 2021) aims to learn how hidden confounders behave over time by using current network observation data and historical information. However, previous works have not bounded confounders leading to high variance estimates when the data-driven approach produces anomaly confounders which have exceeded the impact constraint over treatments and outcomes.

Problem Setup

Estimating Treatment Effects Task

Here we define the problem of estimating treatment effects from irregular time series observations formally: observational data for each patient i at irregular time steps $t_0^i < \cdots < t_{m_i}^i$ for some $m_i \in \mathbb{N}$. We have observed covariates $X^i = [X_{t_0}^i, X_{t_1}^i, \ldots, X_{t_{m_i}}^i] \in \mathcal{X}_t$ and corresponding treatments $A^i = [A_{t_0}^i, A_{t_1}^i, \ldots, A_{t_{m_i}}^i] \in \mathcal{A}_t$, and a_{t_k} is the set of all j possible assigned treatments at timestep t_k. Additionally, we have hidden confounder variables $Z^i = [Z_{t_0}^i, Z_{t_1}^i, \ldots, Z_{t_{m_i}}^i] \in \mathcal{Z}_t$. We omit the patient id i on timestamps unless they are explicitly needed. Combining all hidden confounders, observed covariates, and observed treatments, we define the history before time t_k as $H_{t_k}^i = \{X_{<t_k}^i, A_{<t_k}^i, Z_{<t_k}^i\}$ as the collection of all historical information.

We focus on one-dimensional outcomes $Y^i = [y_{t_0}^i, y_{t_1}^i, \ldots, y_{t_m}^i] \in \mathcal{Y}_t$ and we will be interested in the final expected outcome $\mathbb{E}[Y_{a_t, t_m}^i | H_t^i, X_t^i, A_t^i, Z_t^i]$, given a specified treatment plan a. In this way, we can define the individual treatment effect (ITE) with historical data as $\tau_t^i = \mathbb{E}[Y_{b_t, t_m}^i | H_t^i, X_t^i, A_t^i, Z_t^i] - \mathbb{E}[Y_{a_t, t_m}^i | H_t^i, X_t^i, A_t^i, Z_t^i]$ for two specified treatments a and b. In practice, we rely on assumptions to be able to estimate τ_t^i for any possible treatment plan, which begins at time step t until just before the final patient outcome Y is measured:

Assumption 1. Consistency (Lim, Alaa, and van der Schaar 2018). *If $A_{\geq t} = a_{\geq t}$, then the potential outcomes for following the treatment plan $a_{\geq t}$ is the same as the observed (factual) outcome $Y_{a_{\geq t}} = Y$.*

Assumption 2. Positivity (Overlap) (Imai and Van Dyk 2004). *For any patient, if the probability $P(a_{<t_m}, z_{<t_m}, x_{\leq t_m}) \neq 0$ then the probability of assigning*

treatment: $P(A_{t_m} = a_{t_m} | a_{<t_m}, z_{<t_m}, x_{\leq t_m}) > 0$ *for all a_{t_m}.*

Assumption 1 and Assumption 2 are relatively standard assumptions of causal inference which assume that artificially assigning a treatment has the same impact as if it were naturally assigned and that each treatment has some nonzero probability. Additionally, most previous works in the time series domian make the sequential strong ignorability assumption (Robins and Hernán 2009) that if there are no hidden confounders, for all possible treatments A_t, given the historical observed covariates X_t, we have: $Y_{a_{\geq t_m}} \perp\!\!\!\perp A_{t_m} | A_{<t_m}, X_{<t_m}$. However, this assumption is often untestable due to the presence of hidden confounders in the real-world. Inspired by (Wang and Blei 2019) and (Bica, Alaa, and Van Der Schaar 2020), we assume sequential single strong ignorability in the continuous time setting:

Assumption 3. Sequential single strong ignorability in continuous time setting. *If there exists multi-cause confounders, we have $Y_{a_{\geq t_m}} \perp\!\!\!\perp A_{t_m} | X_{t_m}, H_{<t_m}$, for all $a_{\geq t_m}$ and all j possible assigned treatments.*

Assumption 3 expands the sequential single strong ignorability assumption from (Bica, Alaa, and Van Der Schaar 2020) to the continuous time setting. Thus, only multi cause hidden confounders exist at every time stamp, having a causal effect on the treatment A_t and potential outcome Y_t. One of our goals is to learn representations of hidden confounders under the line of deconfounding works, which aim to eliminate bias, based on the following theorem:

Theorem 1. *If the distribution of the assigned causes $p(a_T)$ can be written as $p(\theta, x_T, z_T, a_T)$, we can obtain sequential ignorable treatment assignment:*

$$Y_{a_{\geq t_m}} \perp\!\!\!\perp A_{t_m} | X_{t_m}, H_{<t_m}, \tag{1}$$

for all $a_{\geq t_m}$ with possible assigned treatments, where θ are the parameters of the causal model.

Thm. 1 is proved by (Bica, Alaa, and Van Der Schaar 2020) and (Hatt and Feuerriegel 2021) in the discrete case. Here, we extend Thm. 1 to the continuous-time setting. Nevertheless, there are still existing challenges in applying the deconfounder framework to longitudinal data in the continuous time setting. After its original publication, (Wang and Blei 2019) has been met with concerns of difficulty in reconstructing confounders in practical applications and the deconfounder assumption itself has been challenged. Towards the necessity of further constraints on the latent confounding, we introduce a frequency-based Lipschitz assumption on the structure of the hidden confounders in Assumption 4.

Assumption 4. Decomposition of time-varying hidden confounders. *The hidden confounders Z_t can be decomposed into high-frequency components Z_t^h and low-frequency Z_t^l with distinguishable frequency gap ω, i.e., $Z_{t_m} = (Z_{t_m}^h, Z_{t_m}^l)$ such that low-frequency confounders have Lipschitz bounded influence and high-frequency confounders are sufficiently covered by proxy variables in X_t.*

In this sense, we combine two existing extensions of TSD under a unifying assumption. Z_t^l contains smooth information

(the trend of the confounding data) bounded by its maximal frequency ω_l. The functional outcomes are then Lipschitz bounded by constant L. Further, its distance and influence from its original value Z_0 will be bounded, reflecting its bounded variation from a static confounder U, as explored in (Hatt and Feuerriegel 2021). Further, the high-frequency components are assumed to have corresponding noisy proxy variables available in the measured covariates X. Consequently, sufficient information about these high-frequency confounders can be derived from the observed proxy variables, as explored in (Kuzmanovic, Hatt, and Feuerriegel 2021). Unified together, our assumption explores a semiparametric assumption enhancing the practicality of applying the deconfounder setup to longitudinal data.

Neural Controlled Differential Equations

Starting from an initial state $u(t_0)$, neural ordinary differential equations (ODE) evolve following a neural network based differential equations. The state at any time t_i is given by integrating an ODE forward in time:

$$\frac{du(t)}{dt} = F(u(t), t; \theta), u(t_i) = u(t_0) + \int_{t_0}^{t_i} \frac{du(t)}{dt} dt, \quad (2)$$

where $F \in \mathcal{F}$, parametrized by θ with $(\mathcal{F}, ||\cdot||)$ a normed vector space and $u(t_0)$ is the initial state. Neural CDEs are a family of continuous time models that explicitly define the latent vector field f_θ by a neural network parameterized by θ, and allow for the dynamics to be modulated by the values of an auxiliary path over time. To constrain the ODE into CDE format, let $\mathbf{H}_t = (H_t^1, H_t^2, \cdots, H_t^n) : t \in [t_0, t_m] \to \mathbb{R}^{n \times m}$ be the m dimensional representation of historical data with all n observed history control paths, the integral be a Riemann-Stieltjes integral and F be a continuous function acting on all control path (Kidger et al. 2020). For continuous time synthetic control, we estimate the latent representation of treatment effect H_t through: $H_t = H_{t_0} + \int_{t_0}^{t} f_\theta(H_s) d\mathbf{H}_s, t \in (t_0, t_m]$.

Lipschitz Bounded Neural Controlled Differential Equations (LipCDE)

To address the treatment effect estimation task from irregular time series observation, we must avoid inference bias caused by hidden confounders. Thus, we propose an approach called Lipschitz bounded neural controlled differential equations (LipCDE). As shown in Figure 1, LipCDE first infers the interrelationship of hidden confounders on treatment by bounding the boundary of hidden confounders via the hidden confounders boundary branch. After that, LipCDE feeds the history trajectories into the synthetic control branch, which utilizes both observed data and hidden confounders to generate the latent representation of each patient. Besides, we re-weight the population of all participating patients and balance the representation via applying a time-varying inverse probability of treatment weighting (IPTW) strategy. Combined with the LSTM layer, the outcome model can get the final estimate of the treatment effect.

Hidden Confounders Boundary Branch

In this section, we focus on how to use Lipschitz regularized convolutional operation to infer the hidden confounders from both high-frequency signals and low-frequency signals of observed data. As shown in Fig 1, the Fourier transform $\mathcal{F}$ on observed data first converts the time-domain signals of history trajectories h_t (Cao et al. 2020, 2021), including covariates and treatments with length N, into the corresponding amplitude and phase at different frequencies. Then, we sort the spectrum so that the spectrum corresponding to low-frequency information is concentrated at the origin after Fourier transform, and high-frequency information is far from the origin and contains rich boundary and detail information. After that, we use Gaussian high-pass filter G_h and Gaussian low-pass filter G_l to get high-frequency components and low-frequency components, respectively:

$$\begin{cases} G_h(h_t) = G_h(\mathcal{F}(h_t)) = 1 - e^{\frac{-D^2(\mathcal{F}(h_t))}{2D_0^2}} \\ G_l(h_t) = G_l(\mathcal{F}(h_t)) = e^{\frac{-D^2(\mathcal{F}(h_t))}{2D_0^2}} \end{cases} \quad (3)$$

The use of spectral-domain analysis enables change detection in certain frequency bands where the influence of trends (low frequency) or daily and seasonal cycles can be considered as time-invariant hidden confounders. The high-frequency components are easily perturbed, which can be treated as noisy proxies. We extract the influence of hidden confounders on the covariates by analyzing the presence of the covariates we extract. After that, both components are fed into convolutional operation:

$$F_c(h_t) = Conv(G_h(h_t)) + Conv(G_l(h_t)) \quad (4)$$

Next, we use the inverse Fourier transform $\mathcal{F}^{-1}$ converts the spectrum information of latent representation back to the time-domain signals. Then, the RNN layer takes the representation $\mathcal{F}^{-1}(F_c(h_t))$ as input and outputs the hidden states h_{hc} of hidden confounders. Note that, after the Fourier transform, time series no longer consider specific timesteps in the spectral domain. In addition, in contrast to directly handling irregular time series as (Ware 1998), we use the processing of the Fourier transform as a mathematical component without considering time intervals, and irregular sampling is enabled in the next component.

To define the boundary of hidden confounders' value interval, following the RNN layer, the confounders encoder uses a Lipschitz bounded linear fully-connected (FC) layer with Lipschitz regularization (Perugachi-Diaz, Tomczak, and Bhulai 2021) to map the output of RNN layer into a hidden embedding, i.e. $z = g(h_{hc}) = W_g h_{hc} + b_g$. The function $g : \mathbb{R}^n \to \mathbb{R}^K$ can be said as L-Lipschitz if there exists an L such that for all $x, y \in \mathbb{R}^n$, we have $||f(x) - f(y)|| \le L||x - y||$ (B'ethune et al. 2021). In this work, we enforce the function g to satisfy the 1-Lipschitz constraint, where g is the linear FC layer. Following spectral normalization of (Gouk et al. 2021):

$$\text{Lip}(g) \le 1, \text{if } ||W_g||_2 \le 1, \quad (5)$$
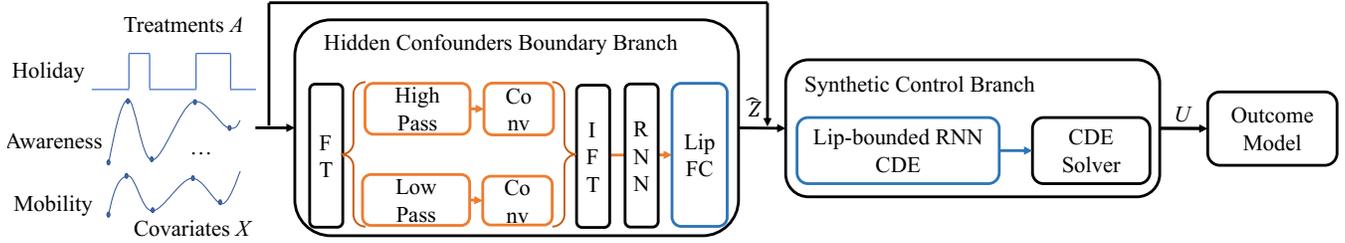
Figure 1: Architecture of LipCDE.

where $|| \cdot ||_2$ is the spectral matrix norm, we enforce the linear weights W_g be at most 1-Lipschitz by having a spectral norm less than one. This constraint ensures that when the observed data is within the normal interval, the inferred hidden confounders satisfy the corresponding bound interval with constant L.

Synthetic Control Branch

Since the neural ordinary differential equations(ODE) family is effective in continuous time problems, we use neural CDE to estimate latent factors and treatment effects. Inspired by (Bellot and Van Der Schaar 2021), let $u_t := g_\eta(H_t) = g_\eta([x_t, a_t, \hat{z}_t, H_{t-1}])$, where $g_\eta : \mathbb{R}^{n \times m} \to \mathbb{R}^{l \times m}$ is a set of functions that embeds the historical data into a l-dimensional latent state. Let f be a neural network parameterizing the latent vector field. To apply Lipschitz constraint on f, following (Erichson et al. 2020), we define f as a continuous time Lipschitz RNN:

$$f(h, t) = A_R h + \sigma(W_R h + U u(s) + b), \quad (6)$$

where hidden-to-hidden matrices A_R and W_R are trainable matrices and nonlinearity $\sigma(\cdot)$ is an 1-Lipschitz function. Now $\dot{f} = \frac{\partial f(t)}{\partial t}$ is the time derivative and f considers both controlling the history path of observed data and the hidden state of RNN. A latent path can be expressed as the solution to a controlled differential equation of the form:

$$u_t = u_{t_0} + \int_{t_0}^{t} f(u_s, s) \, d\mathbf{H}_s^0, \quad t \in (t_0, t_m] \quad (7)$$

In that way, we can directly utilize adjoint methods (Chen et al. 2018) of CDEs to enable computing the gradient with a dynamic causal relationship between historical information controlled by $\mathbf{H}$ and outcomes. For each estimate of f_θ and g_η the forward latent trajectory in time that these functions defined through (7) can be computed using any numerical ODE solver as those equations continuously incorporate incoming interventions, without interrupting the differential equation:

$$\hat{u}_{t_1}, \dots, \hat{u}_{t_k} = \text{ODESolve}(f_\theta, u_{t_0}, \mathbf{H}_{t_1}, \dots, \mathbf{H}_{t_k}) \quad (8)$$

Outcome Model

After sampling the latent representation $U_t = (\hat{u}_{t_1}, \dots, \hat{u}_{t_k})$ of historical trajectories on each patient, we use the outcome model to estimate the treatment effect. To adjust the treatment assignment and get the final estimates, we first re-weight the population via an RNN model, which can handle time-varying treatment assignment (Lim, Alaa, and van der Schaar 2018), to estimate the propensity scores and IPTW of each dynamic time steps. After that, we use two stacked LSTM layers to decode the padded hidden sequence of irregular inputs. Then we use a linear fully-connected layer mapping the output of the LSTM layer into an unbiased estimated treatment response over time. For the loss function part, we weight each patient via the generated score of IPTW, w^i, and use the mean squared error (MSE) function as our target loss function: $L = \frac{1}{N} \sum_{i=1}^{N} w^i(\hat{y}_{t_{m+1}}^i - y_{t_{m+1}}^i)^2$.

Empirically, the identifiability can be assessed on the synthetic data via sample hidden confounders Z_t repeatedly to evaluate the uncertainty of the outcome model estimates. However, identifiability might not be guaranteed under the framework of deconfounding in the completely general case (D'Amour 2019; Ogburn, Shpitser, and Tchetgen 2020). Previous works find that the estimates may have a high variance when the treatment effects are non-identifiable (Bica, Alaa, and Van Der Schaar 2020; Hatt and Feuerriegel 2021; Kuzmanovic, Hatt, and Feuerriegel 2021). To achieve the goal of identifiability and obtain unbiased ITE estimates, (Hatt and Feuerriegel 2021) introduces the assumption of *Time-Invariant Unobserved Confounding*, which requires the hidden confounders are invariant for different timestamps, and (Kuzmanovic, Hatt, and Feuerriegel 2021) claim that we can learn the hidden embedding to make *Sequential Strong Ignorability* assumption hold via the observed noised proxies. Thus, the greater identifiability of our work follows both (Hatt and Feuerriegel 2021) and (Kuzmanovic, Hatt, and Feuerriegel 2021) as it utilizes both time-invariant hidden confounders from low-frequency components and dynamic noisy proxies from the high-frequency component of the observed data simultaneously in practice.

Experiments

Experiments Setting

In this section, we estimate the treatment effects for each time step by one-step ahead predictions on both synthetic dataset and real-world datasets including MIMIC-III (Johnson et al. 2016) dataset and COVID-19 (Steiger, Mußgnug, and Kroll 2020) dataset. Hidden confounders in such real-world datasets is present as variables not included in the records. However, for real-world data, it is untestable to estimate the oracle treatment responses and we only evaluate the factual treatment effects.

(a) RMSE results on treatment effects

(b) RMSE results on counterfactual treatment effects
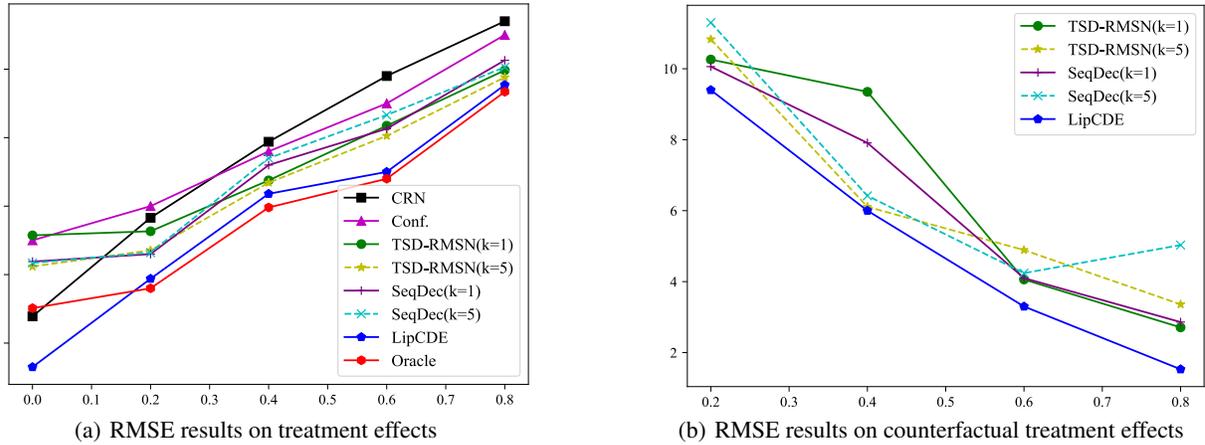
Figure 2: Results on synthetic data. The x-axis of each graph is the confounding degree and the y-axis is RMSE($\times 100\%$).

Baselines. LipCDE is evaluated by examining the degree of control it has over hidden confounders. The baselines used in these experiments are: **Oracle**, which estimates ITE with simulated (oracle) confounders; **Conf. (No-hidden)**, which assumes no hidden confounders and can make it clear how hidden confounders here impact the performance of treatment effect prediction models; **CRN** (Bica et al. 2020), which introduces a sequence-to-sequence counterfactual recurrent network to estimate treatment effects and utilizes domain adversarial training to handle the bias from time-varying confounders; **TSD** (Bica, Alaa, and Van Der Schaar 2020), which leverages the assignment of multiple treatments over time to enable the estimation of treatment effects in the presence of multi-cause hidden confounders; **DTA** (Kuzmanovic, Hatt, and Feuerriegel 2021), which combines a LSTM autoencoder with a causal regularization penalty to learn dynamic noisy proxies and render the potential outcomes and treatment assignment conditionally independent; **SeqDec** (Hatt and Feuerriegel 2021) , which utilizes a Gaussian process latent variable model to infer substitutes for the hidden confounders; **OriCDE** (Bellot and Van Der Schaar 2021), which can estimate ITE explicitly using the formalism of linear controlled differential equations. Except OriCDE, all baselines share the same design of the outcome model, i.e. *MSM* (Robins, Hernán, and Brumback 2000), which uses inverse probability of treatment weighting (IPTW) to adjust for the time-dependent confounding bias by linear regression and then constructs a pseudo-population to compute final outcome, and *RMSN* (Lim, Alaa, and van der Schaar 2018), which estimates IPTW using RNNs instead of logistic regressions. OriCDE and LipCDE use the outcome model introduced in previous section.

Estimating Treatment Effects Experiments

Synthetic experiments. For the synthetic dataset, in addition to estimating factual treatment responses, we will also perturb the inputs to quantify how accurate counterfactual relationships are captured by LipCDE. Following TSD, we have $T = 30$ max time steps and $N = 5000$ patient trajecto-

ries, where each patient has $p = 5$ observed covariates and different treatments. We vary the confounding degree parameter γ to produce a varying amount of hidden confounders. Factual results use the outcome results corresponding to the real-world treatment we simulate. For the counterfactual estimations, we set all the treatments to 0 at the timestamp interval of $[\frac{l_i}{2}, l_i]$, where l_i is the sequential length of patient i, and get the outcome of the counterfactual world. As shown on Figure 2, for the factual treatment effects results, methods considering hidden confounders are generally better than the models without the hidden confounders (CRN, Conf.). Note that, LipCDE achieves better results on all different levels of confounders and its outcome is closest to the estimates obtained using simulated (oracle) confounders, which means LipCDE can yield less biased estimates compared with other baselines. In addition, LipCDE remains stable and becomes closer to the simulated (oracle) confounders baseline when we increase the degree of confounders influence, which indicates that our model can effectively constrain the influence boundary of hidden confounders based on observed data. For the counterfactual path results, we interestingly observe that the RMSE decreases as the confounding degree increases. The reason is that when the degree increase, Z_t gets easier to handle with fixed treatment plans referring to the data generation method. Besides, LipCDE still performs better than the current baselines in the counterfactual world, indicating the stability of LipCDE for hidden confounder's reasoning and the validity of the estimation.

Real-world experiments on MIMIC-III & COVID-19. real-world data allow us to demonstrate LipCDE has strong scalability and interpretability in real-world applications. MIMIC-III dataset contains 5000 patient records with 3 treatments, 20 covariates of patients and 2 outcomes including blood pressure (Blo. pre.), and oxygen saturation (Oxy. sat.). The COVID-19 dataset contains 401 German districts over the period of 15 February to 8 July 2020. We extract 10 time-varying covariates and 2 treatments with 2 outcomes, 'active cases', in each district. The results in Table 1 show that LipCDE outperforms existing baselines in all cases. By

Outcome Model	-	MSM (RMSE%)			RMSN (RMSE%)				Ours (RMSE%)	
Methods	CRN	Conf.	DTA	TSD	Conf.	DTA	TSD	SeqDec	OriCDE	LipCDE
Blo. pre.	12.43	14.54	13.31	13.57	14.46	18.33	12.11	13.74	10.55	**9.19**
Oxy. sat.	4.17	4.72	4.65	4.33	4.22	4.21	4.25	4.19	4.24	**4.15**
COVID-19	-	15.10	13.93	13.07	11.48	13.52	11.08	11.43	11.36	**7.56**

Table 1: Results for real-world data (MIMIC-III and COVID-19) experiments. Lower is better.

Degree	MR	Conf.	TSD	SeqDec	LipCDE	MR	Conf.	TSD	SeqDec	LipCDE
0		3.43	2.83	2.43	**1.19**		3.32	2.84	3.19	**2.29**
0.2	15%	3.47	2.84	2.69	**2.6**	30%	4.66	3.65	2.95	**2.62**
0.4		3.45	3.67	3.7	**3.39**		4.19	4.06	3.89	**3.61**

Table 2: Irregular data with missing value

modeling the dependence of the assigned treatment for each patient, LipCDE is able to infer latent variables and make orderly use of the causal relationship between latent variables and observed data. This result is consistent with what we have seen in the simulated dataset. Specifically, the average RMSE on MIMIC-III's blood pressure outcome and COVID-19 datasets is decreased by 28.7% and 32.3% over TSD and SeqConf respectively. Besides, the small increase in oxygen saturation is thought to be due to the fact that oxygen saturation itself is not dependent on current covariates and is less influenced by treatment. Although these results on real data require further validation by physicians, they demonstrate the potential of the method to be applied in real medical scenarios.
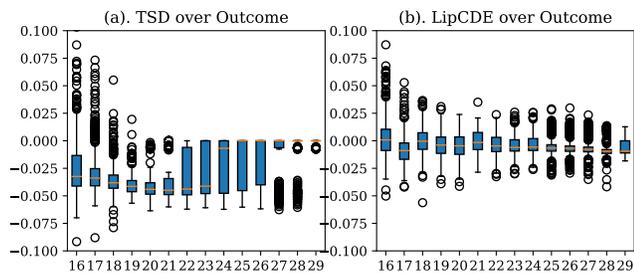
Figure 3: Analysis of the outcome on synthetic data's counterfactual path. Comparing with baseline models, LipCDE can estimate treatment effects with lower variance

Analysis

Irregular time series with missing values. We emphasize that our model is suitable for irregular time series sampling. Therefore, we remove randomly 15% and 30% of the aligned synthetic data with different confounding degrees, independently for each unit. Except for CDE-based methods, all the baselines require some form of prior interpolation. Results shown in Table 2 demonstrate that our model achieves a comparable performance with irregularly aligned data. Note that, comparing with SeqDec which only models irregular samples via an indirect simple multivariate Gaussian distribution, LipCDE shows the ability of handling continuous time setting by utilizing the CDE module.

Analysis on bounded hidden confounders. We perform the analysis using simulated datasets and evaluate the hidden confounder's quality on LipCDE with TSD and SeqDec. As shown on Figure 3, LipCDE can achieve better estimate results with lower variance compared with the previous strong baseline. Further, we find that TSD can induce highly confident posterior distributions with lower bounds of the hidden confounders, which can yield highly confident biased predictions (Zheng, D'Amour, and Franks 2021). The seqDec has more discrete points and no obvious boundary, which also leads to the degradation of the model performance. LipCDE controls the data distribution of hidden confounders more accurately by filtering the convolutional neural network and Lipschitz regularization, which has higher similarity to the originally hidden confounder compared with other baselines.

Conclusion

In this paper, we proposed the Lipschitz-bounded neural controlled differential equation (LipCDE), a novel neural network that utilizes hidden confounders for estimating treatment effect in the case of irregular time series observations. For one thing, it uses the performance of time-varying observations in the frequency domain to infer the hidden confounders under Lipschitz regularization. For another thing, a well-designed CDE explicitly models the combinational latent path of observed time series, which can effectively capture underlying temporal dynamics and intervention effects. With experimental results on synthetic and real datasets, we demonstrate the effectiveness of LipCDE in reducing bias in the task of estimating treatment effects.

Acknowledgements

References

Abadie, A.; Diamond, A.; and Hainmueller, J. 2010. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, 105(490): 493–505.

Araujo, A.; Negrevergne, B.; Chevaleyre, Y.; and Atif, J. 2021. On Lipschitz Regularization of Convolutional Layers using Toeplitz Matrix Theory. *Thirty-Fifth AAAI Conference on Artificial Intelligence*.

Bahadori, M. T.; and Heckerman, D. 2021. Debiasing Concept-based Explanations with Causal Analysis. In *International Conference on Learning Representations*.

Bellot, A.; and Van Der Schaar, M. 2021. Policy analysis using synthetic controls in continuous-time. In *International Conference on Machine Learning*, 759–768. PMLR.

Berrevoets, J.; Jordon, J.; Bica, I.; Gimson, A.; and van der Schaar, M. 2020. OrganITE: Optimal transplant donor organ offering using an individual treatment effect. *https://proceedings. neurips. cc/paper/2020*, 33.

B'ethune, L.; Gonz'alez-Sanz, A.; Mamalet, F.; and Serrurier, M. 2021. The Many Faces of 1-Lipschitz Neural Networks. *ArXiv*, abs/2104.05097.

Bica, I.; Alaa, A.; and Van Der Schaar, M. 2020. Time series deconfounder: Estimating treatment effects over time in the presence of hidden confounders. In *International Conference on Machine Learning*, 884–895. PMLR.

Bica, I.; Alaa, A. M.; Jordon, J.; and van der Schaar, M. 2020. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *International Conference on Learning Representations*.

Brodersen, K. H.; Gallusser, F.; Koehler, J.; Remy, N.; and Scott, S. L. 2015. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1): 247–274.

Cao, D.; Li, J.; Ma, H.; and Tomizuka, M. 2021. Spectral temporal graph neural network for trajectory prediction. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 1839–1845. IEEE.

Cao, D.; Wang, Y.; Duan, J.; Zhang, C.; Zhu, X.; Huang, C.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; et al. 2020. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems*, 33: 17766–17778.

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. 2018. Neural ordinary differential equations. *arXiv preprint arXiv:1806.07366*.

Curth, A.; and van der Schaar, M. 2021. On Inductive Biases for Heterogeneous Treatment Effect Estimation. *arXiv preprint arXiv:2106.03765*.

D'Acunto, G.; Bajardi, P.; Bonchi, F.; and De Francisci Morales, G. 2021. The evolving causal structure of equity risk factors. In *Proceedings of the Second ACM International Conference on AI in Finance*, 1–8.

D'Amour, A. 2019. On Multi-Cause Causal Inference with Unobserved Confounding: Counterexamples, Impossibility, and Alternatives. *ArXiv*, abs/1902.10286.

De Brouwer, E.; Gonzalez, J.; and Hyland, S. 2022. Predicting the impact of treatments over time with uncertainty aware neural differential equations. In *International Conference on Artificial Intelligence and Statistics*, 4705–4722. PMLR.

Erichson, N. B.; Azencot, O.; Queiruga, A.; Hodgkinson, L.; and Mahoney, M. W. 2020. Lipschitz recurrent neural networks. *arXiv preprint arXiv:2006.12070*.

Fröhlich, F.; Loos, C.; and Hasenauer, J. 2019. Scalable inference of ordinary differential equation models of biochemical processes. *Gene Regulatory Networks*, 385–422.

Gao, T.; Subramanian, D.; Bhattacharjya, D.; Shou, X.; Mattei, N.; and Bennett, K. P. 2021. Causal inference for event pairs in multivariate point processes. *Advances in Neural Information Processing Systems*, 34: 17311–17324.

Gouk, H.; Frank, E.; Pfahringer, B.; and Cree, M. J. 2021. Regularisation of neural networks by enforcing lipschitz continuity. *Machine Learning*, 110(2): 393–416.

Guo, R.; Li, J.; and Liu, H. 2020a. Counterfactual evaluation of treatment assignment functions with networked observational data. In *Proceedings of the 2020 SIAM International Conference on Data Mining*, 271–279. SIAM.

Guo, R.; Li, J.; and Liu, H. 2020b. Learning individual causal effects from networked observational data. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 232–240.

Hatt, T.; and Feuerriegel, S. 2021. Sequential Deconfounding for Causal Inference with Unobserved Confounders. *arXiv preprint arXiv:2104.09323*.

Hernán, M. A.; and Robins, J. M. 2010. *Causal inference: What If*. CRC Boca Raton, FL.

Imai, K.; and Van Dyk, D. A. 2004. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99(467): 854–866.

Johansson, F.; Shalit, U.; and Sontag, D. 2016. Learning representations for counterfactual inference. In *International conference on machine learning*, 3020–3029. PMLR.

Johnson, A. E.; Pollard, T. J.; Shen, L.; Li-Wei, H. L.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Celi, L. A.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1): 1–9.

Kallus, N.; Mao, X.; and Udell, M. 2018. Causal inference with noisy and missing covariates via matrix factorization. *arXiv preprint arXiv:1806.00811*.

Kidger, P.; Morrill, J.; Foster, J.; and Lyons, T. 2020. Neural controlled differential equations for irregular time series. *arXiv preprint arXiv:2005.08926*.

Kuroki, M.; and Pearl, J. 2014. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2): 423–437.

Kuzmanovic, M.; Hatt, T.; and Feuerriegel, S. 2021. Deconfounding Temporal Autoencoder: estimating treatment effects over time using noisy proxies. In *Machine Learning for Health*, 143–155. PMLR.

Langley, P. 2000. Crafting Papers on Machine Learning. In Langley, P., ed., *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, 1207–1216. Stanford, CA: Morgan Kaufmann.

Lim, B.; Alaa, A.; and van der Schaar, M. 2018. Forecasting Treatment Responses Over Time Using Recurrent Marginal Structural Networks. *NeurIPS*, 18: 7483–7493.

Liu, R.; Yin, C.; and Zhang, P. 2020. Estimating individual treatment effects with time-varying confounders. In *2020 IEEE International Conference on Data Mining (ICDM)*, 382–391. IEEE.

Ma, J.; Guo, R.; Chen, C.; Zhang, A.; and Li, J. 2021. Deconfounding with Networked Observational Data in a Dynamic Environment. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 166–174.

Mandel, T.; Liu, Y.-E.; Levine, S.; Brunskill, E.; and Popovic, Z. 2014. Offline policy evaluation across representations with applications to educational games. In *AAMAS*, volume 1077.

Mastakouri, A. A.; Schölkopf, B.; and Janzing, D. 2021. Necessary and sufficient conditions for causal feature selection in time series with latent common causes. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, 7502–7511. PMLR.

Ogburn, E. L.; Shpitser, I.; and Tchetgen, E. J. T. 2020. Counterexamples to" the blessings of multiple causes" by wang and blei. *arXiv preprint arXiv:2001.06555*.

Pearl, J. 2012. On measurement bias in causal inference. *arXiv preprint arXiv:1203.3504*.

Pearl, J.; et al. 2000. Causality: Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, 19.

Perugachi-Diaz, Y.; Tomczak, J. M.; and Bhulai, S. 2021. Invertible DenseNets with Concatenated LipSwish. In Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*.

Robins, J. M.; and Hernán, M. A. 2009. Estimation of the causal effects of time-varying exposures. *Longitudinal data analysis*, 553: 599.

Robins, J. M.; Hernán, M. A.; and Brumback, B. 2000. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, 11(5): 551.

Roy, J.; Lum, K. J.; and Daniels, M. J. 2017. A Bayesian nonparametric approach to marginal structural models for point treatments and a continuous or survival outcome. *Biostatistics*, 18(1): 32–47.

Schulam, P.; and Saria, S. 2017. Reliable decision support using counterfactual models. *Advances in Neural Information Processing Systems*, 30: 1697–1708.

Seedat, N.; Imrie, F.; Bellot, A.; Qian, Z.; and van der Schaar, M. 2022. Continuous-Time Modeling of Counterfactual Outcomes Using Neural Controlled Differential Equations. *arXiv preprint arXiv:2206.08311*.

Shalit, U.; Johansson, F. D.; and Sontag, D. 2017. Estimating individual treatment effect: generalization bounds and algorithms. In *International Conference on Machine Learning*, 3076–3085. PMLR.

Simpson, E. H. 1951. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2): 238–241.

Soleimani, H.; Subbaswamy, A.; and Saria, S. 2017. Treatment-response models for counterfactual reasoning with continuous-time, continuous-valued interventions. *arXiv preprint arXiv:1704.02038*.

Steiger, E.; Mußgnug, T.; and Kroll, L. E. 2020. Causal analysis of COVID-19 observational data in German districts reveals effects of mobility, awareness, and temperature. *medRxiv*, 2020–07.

Veitch, V.; Sridhar, D.; and Blei, D. 2020. Adapting text embeddings for causal inference. In *Conference on Uncertainty in Artificial Intelligence*, 919–928. PMLR.

Wang, Y.; and Blei, D. M. 2019. The blessings of multiple causes. *Journal of the American Statistical Association*, 114(528): 1574–1596.

Ware, A. F. 1998. Fast approximate Fourier transforms for irregularly spaced data. *SIAM review*, 40(4): 838–856.

Zhang, Y.; Cao, D.; and Liu, Y. 2022. Counterfactual Neural Temporal Point Process for Estimating Causal Influence of Misinformation on Social Media. In *Advances in Neural Information Processing Systems*.

Zheng, J.; D'Amour, A.; and Franks, A. 2021. Bayesian Inference and Partial Identification in Multi-Treatment Causal Inference with Unobserved Confounding. *arXiv preprint arXiv:2111.07973*.

Zibelman, M.; Ramamurthy, C.; and Plimack, E. R. 2016. Emerging role of immunotherapy in urothelial carcinoma—advanced disease. In *Urologic Oncology: Seminars and Original Investigations*, volume 34, 538–547. Elsevier.